

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 2, 2001, 04:55:10 ; Search time 67.53 Seconds
(without alignments)
4132.158 Million cell updates/sec

Title: US-09-537-654-1
Perfect score: 1474
Sequence: 1 tcgaccacgcgcgcgcact.....aaaaaaaaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCRTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfilea1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.8	3.6	3777	3 US-09-121-321-15	Sequence 15, Appl
2	53.8	3.6	3777	4 US-08-933-803A-15	Sequence 15, Appl
3	49.8	3.4	1641	1 US-08-300-903A-8	Sequence 8, Appl
4	48.4	3.3	1824	3 US-08-606-505B-1	Sequence 1, Appl
5	48.4	3.3	1824	4 US-09-616-990-1	Sequence 1, Appl
6	47.8	3.2	958	2 US-08-757-046A-5	Sequence 5, Appl
7	47.8	3.2	958	3 US-09-447-208-5	Sequence 5, Appl
8	47.8	3.2	958	4 US-09-135-988-5	Sequence 5, Appl
9	47.8	3.2	958	4 US-09-277-716-5	Sequence 5, Appl
10	47.8	3.2	958	4 US-08-597-274A-5	Sequence 5, Appl
11	47.8	3.2	536	1 US-08-341-568-1	Sequence 1, Appl
12	47.8	3.2	536	2 US-08-811-020-1	Sequence 1, Appl
13	46.6	3.2	2852	3 US-09-027-137-2	Sequence 2, Appl
14	46.6	3.2	1386	2 US-08-897-340-1	Sequence 1, Appl
15	46.6	3.2	1386	3 US-09-252-329-1	Sequence 1, Appl
16	46.4	3.1	2447	2 US-09-014-969-14	Sequence 14, Appl
17	46.2	3.1	2296	4 US-08-496-841C-137	Sequence 13, Appl
18	46.2	3.1	4765	3 US-08-936-135-21	Sequence 21, Appl
19	46.2	3.1	4780	3 US-08-936-135-23	Sequence 23, Appl
20	46.2	3.1	4780	3 US-08-936-135-23	Sequence 23, Appl
21	46.2	3.1	375	3 US-08-946-026-23	Sequence 3, Appl
22	46.2	3.1	748	1 US-08-361-467B-3	Sequence 3, Appl
23	45.8	3.1	1813	5 PCN-US94-12883-3	Sequence 3, Appl
24	45.8	3.1	1813	5 PCN-US94-12883-3	Sequence 3, Appl
25	45.8	3.1	2345	3 US-08-955-937A-1	Sequence 1, Appl
26	45.4	3.1	2345	3 US-08-955-937A-1	Sequence 1, Appl
27	45.2	3.1	882	4 US-08-909-965C-9	Sequence 9, Appl
			1858	2 US-08-909-965C-11	Sequence 11, Appl

28	45	3.1	1781	1 US-08-314-615-2	Sequence 2, Appl
29	45	3.1	1781	1 US-08-314-615-2	Sequence 2, Appl
30	45	3.1	1781	1 US-08-433-010-2	Sequence 2, Appl
31	45	3.1	1781	1 US-08-433-010-2	Sequence 2, Appl
32	45	3.1	1781	1 US-08-245-295-4	Sequence 4, Appl
33	45	3.1	1781	1 US-08-481-130-4	Sequence 4, Appl
34	45	3.1	1781	1 US-08-656-984A-4	Sequence 4, Appl
35	45	3.1	1781	1 US-08-482-882-2	Sequence 4, Appl
36	45	3.1	1781	1 US-08-483-604-4	Sequence 4, Appl
37	45	3.1	1781	1 US-08-483-389-2	Sequence 4, Appl
38	45	3.1	1781	2 US-08-487-113D-2	Sequence 2, Appl
39	45	3.1	1781	2 US-08-487-595-4	Sequence 2, Appl
40	45	3.1	1781	2 US-08-473-503-2	Sequence 2, Appl
41	45	3.1	1781	2 US-08-483-932-2	Sequence 2, Appl
42	45	3.1	1781	2 US-08-720-420A-2	Sequence 2, Appl
43	45	3.1	1781	3 US-08-714-017-2	Sequence 2, Appl
44	45	3.1	1781	3 US-08-863-790-2	Sequence 2, Appl
45	45	3.1	1781	3 US-08-475-680-2	Sequence 2, Appl
			1781	3 US-08-296-749-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-121-321-15
; Sequence 15, Application US/09121321
; Patent No. 6090763
; GENERAL INFORMATION:
; APPLICANT: Saiga, Akihiko
; APPLICANT: Orita, Satoshi
; APPLICANT: Igarashi, Hisanaga
; APPLICANT: Okumura, Kouichi
; APPLICANT: Sakaguchi, Gaku
; TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
; TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/121,321
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/933,803
; FILING DATE: 19-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: SHGN-12C1P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3777 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 139..2151
; US-09-121-321-15

RESULT 5
US-09-616-990-1
; Sequence 1, Application US/09616990
; Patent No. 6232109
; GENERAL INFORMATION:
; APPLICANT: KIRUCHI, Yasuhiro
; KIYOKAWA, Shigeto
;

RESULT 6
US-08-757-046A-5
; Sequence 5, Application US/08757046A
; Patent No. 5876995
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF
; NUMBER OF SOUNDENCES: 14 MANUFACTURE

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,046A
FILING DATE: 11-25-96
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/597,274
FILING DATE: 02-06-96
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6680-105B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 115...702
OTHER INFORMATION: apoasequorlin-encoding gene
PUBLICATION INFORMATION:
AUTHORS: Inouye et al.
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 82
PAGES: 3154-3158
DATE: (1985)
DOCUMENT NUMBER: PATENT NO.: 5,093,240
US-08-757-046A-5

Query Match 3.2%, Score 47.8; DB 2; Length 958;
Best Local Similarity 67.7%; Pred. No. 0.00047;
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1376 tagtttttgcgtcgaacatctggcccatcgtaacatcgaagcctataatagcggg 1435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 856 TGTGTTTTTAATCAACAGAACTTCAATCGAAAAAGTAAAAAAGTAAAAAAGTAAAAA 915
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1436 caaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 1474
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 916 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954

RESULT 7
US-09-447-208-5
; Sequence 5, Application US/09447208
; Patent No. 6113886
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
; NUMBER OF SEQUENCES: 14

```

```

CORRESPONDENCE ADDRESS:
ADDRESSSEE: Heller Ehrman White & McCauliffe
STREET: 4250 Executive Square, 7th Floor
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/447,208
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0909/135,988
FILING DATE: 08-17-98
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/757,046
FILING DATE: 11-25-96
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/597,274
FILING DATE: 02-06-96
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24727-105C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-450-8400
TELEFAX: 619-450-8499
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 115...702
OTHER INFORMATION: apoaecuatorin-encoding gene
PUBLICATION INFORMATION:
PUBLICATION INFORMATION: PATENT NO.: 5,093,240
AUTHORS: Imouye et al.
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 82
PAGES: 3154-3158
DATE: (1985)
US-09-447-208-5

Query Match          3.2%; Score 47.8; DB 3; Length 958;
Best Local Similarity 67.7%; Pred.No. 0.00047;
Matches    67; Conservative    0; Mismatches   32; Indels      0; Gaps      0;

QY 1376 taatgttttctgatacgaacatctggccctcgtaacttcagtaaaggccataatagcg99 1435
DB 856 ttgtttttttaacaacaaaccttacaaaatcgaaaaagtaaaaaaaaaaaaaaaaaaaaaa 915
QY 1436 caaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474
DB 916 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 954

```



```

Db          916 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954
|||||
RESULT      9
US-09-277-716-5
; Sequence 5, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: TITLEFEASAS, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 958
; TYPE: DNA
; ORGANISM: Aequorea (luminescent jellyfish)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (115)..(702)
; FEATURE:
; OTHER INFORMATION: Apoaequorin-encoding gene
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5,093,240
; PATENT FILING DATE: 1987-10-08
; PUBLICATION DATE: 1992-03-03
; PUBLICATION INFORMATION:
; AUTHORS: Inouye, S.
; TITLE: Cloning and sequence analysis of cDNA for the luminescent protein
; JOURNAL: Proc. Natl. Acad. Sci. USA
; VOLUME: 82(10)
; PAGES: 3154-3158
; DATE: 1985-05
US-09-277-716-5

Query Match      3.2%; Score 47.8; DB 4; Length 958;
Best Local Similarity 67.7%; Pred. No. 0.00047;
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1376 taggtttgtatctgaacatctgcccattcgtacatcagtaagcctataatagcggg 1435
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 ttgtttttaaataacacgaacttacaatctgaaataagtaaaaaaataaaaaaa 915
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1436 caaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 1474
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 916 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 954

RESULT      10
US-08-597-274A-5
; Sequence 5, Application US/08597274A
; Patent No. 6247995
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT NOVELTY ITEMS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA

```

ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/597,274A
FILING DATE: 02/06/96
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6680-105
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 115...702
OTHER INFORMATION: apoaeguorIn-encoding gene
PUBLICATION INFORMATION:
DOCUMENT NUMBER: 5,093,240
AUTHORS: Inouye et al.
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 82
PAGES: 3154-3158
DATE: (1985)
US-08-597-274A-5

Query Match 3.2%; Score 47.8; DB 4; Length 958;
Best Local Similarity 67.7%; Pred. No. 0.00047;
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1376 tagtgtttgtatcgtgacatctgcccattcagtaagccataataagcgg 1435
DB 856 ttgttttttttTCAACACACTTCAAAATCGAAAAAGTAAAAA 915
QY 1436 caaaaaa 1474
DB 916 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954

RESULT 11
US-08-341-568-1
Sequence 1, Application US/08341568
Patent No. 5651021
GENERAL INFORMATION:
APPLICANT: Buchert, Johanna
APPLICANT: Slika-aho, Matti
APPLICANT: Viikari, Liisa
APPLICANT: Penttila, Merja
APPLICANT: Saloheimo, Anu
APPLICANT: Marjatta, Rannu
TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching
TITLE OF INVENTION: pulps

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: PO Box 747
CITY: Falls Church
STATE: VA
COUNTRY: US
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/341,568
FILING DATE: 22-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 365-262P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
STRAIN: QM9414
US-08-341-568-1

Query Match 3.2%; Score 47; DB 1; Length 536;
Best Local Similarity 64.0%; Pred. No. 0.00057;
Matches 71; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1364 aacaagctaataagttgtatctgacatctgcccattcagtaagcc 1423
DB 411 AACCTTTTACATTGATTACTGCGTATGAACAAGTTGTAGGCTGTAACGAAAAA 470
QY 1424 tataatagcggcAaaaaa 1474
DB 471 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 521

RESULT 12
US-08-911-020-1
Sequence 1, Application US/08911020
Patent No. 5854047
GENERAL INFORMATION:
APPLICANT: Buchert, Johanna
APPLICANT: Slika-aho, Matti
APPLICANT: Viikari, Liisa
APPLICANT: Penttila, Merja
APPLICANT: Saloheimo, Anu
APPLICANT: Marjatta, Rannu
TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching
TITLE OF INVENTION: lignocellulosic pulps
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: PO Box 747
CITY: Falls Church
STATE: VA
COUNTRY: US
ZIP: 22040-0747

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,020
FILING DATE: 13-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/341,568
FILING DATE: 22-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 365-262P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
STRAIN: QM9414
US-08-911-020-1

Query Match 3.2%; Score 47; DB 2; Length 536;
Best Local Similarity 64.0%; Pred. No. 0.00057;
Matches 71; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1364 aaccagcctaataagttgttctgatacgtcgccatcgtaacagtcg 1423
DB 411 AACCTCTTACATGATGATGCTGATGACAAAGGTGTGGGTAGGTAAAGAAAAA 470
QY 1424 tatatagcggcgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474
DB 471 AA 521

RESULT 13
US-09-027-137-2
Sequence 2, Application US/09027137
Patent No. 6013450
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Yue, Henry
TITLE OF INVENTION: CAF1-RELATED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/027,137
FILING DATE: Filed Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0476 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2852 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNOT16
CLONE: 2229466
US-09-027-137-2

Query Match 3.2%; Score 47; DB 3; Length 2852;
Best Local Similarity 59.3%; Pred. No. 0.0014;
Matches 80; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1340 ttaagagagaagtaataacacagacagcctaataatagtcggcgcaaaaaaaaaaaaaaaaaaaaaa 1459
DB 2432 TTAAAGATGCTTGTAAATCACTATTTATTTTAAATTTTGTAAATTAAGATTTCTTTTAA 2491
QY 1400 gccatcgtacatcagtaagcctaataatagtcggcgcaaaaaaaaaaaaaaaaaaaaaa 1459
DB 2492 ACCACTGCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2551
QY 1460 aaaaaaaaaaaaaa 1474
DB 2552 AAAAAAAAAAAAAAAAAAAAAA 2566

RESULT 14
US-08-897-340-1
Sequence 1, Application US/08897340
Patent No. 5955306
GENERAL INFORMATION:
APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
TITLE OF INVENTION: Weight Control Pathway Genes and Uses
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,340
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/715,032
FILING DATE: 17-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Silverl, Jean M.
REGISTRATION NUMBER: 39,030
REFERENCE/DOCKET NUMBER: MNI-005CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1386 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-897-340-1

Query Match 3.2%; Score 46.6; DB 2; Length 1386;
 Best Local Similarity 64.6%; Pred. No. 0.0012;
 Matches 64; Conservative 3; Mismatches 32; Indels 0; Gaps 0;

QY 1376 tagtggtttctgaacatctggccatctacatcagtaagcctataatagcgg 1435
 DB 1267 TTGGTTTTCATTTTGTGCTTCCTTCCTAGATCCAGAAATTAAGTTAAGRGRAGSGC 1326

QY 1436 caaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474
 DB 1327 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1365

RESULT 15
 US-09-252-329-1

; Sequence 1, Application US/09252329

; Patent No. 6147192

; GENERAL INFORMATION:

; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.

; TITLE OF INVENTION: Weight Control Pathway Genes and Uses

; NUMBER OF SEQUENCES: 36

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHYE & COCKFIELD, LLP

; STREET: 28 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/252,329

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/897,340

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Silveri, Jean M.

; REGISTRATION NUMBER: 39,030

; REFERENCE/DOCKET NUMBER: MNT-005CP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1386 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-09-252-329-1

Query Match 3.2%; Score 46.6; DB 3; Length 1386;
 Best Local Similarity 64.6%; Pred. No. 0.0012;
 Matches 64; Conservative 3; Mismatches 32; Indels 0; Gaps 0;

QY 1376 tagtggtttctgaacatctggccatctacatcagtaagcctataatagcgg 1435

Search completed: October 2, 2001, 05:48:56
 Job Time: 3226 sec

DB 1267 TTGGTTTTCATTTTGTGCTTCCTTCCTAGATCCAGAAATTAAGTTAAGRGRAGSGC 1326

QY 1436 caaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474
 DB 1327 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1365

Tue Oct 2 11:16:25 2001

us-09-537-654-1.rni

Page 9

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time 20.7 Seconds
(without alignments)
861.036 Million cell updates/sec

Title: US-09-537-654-2

Perfect score: 1530
Sequence: 1 MDDSGSNGNGPOGXVSGAQ.....VTGKGRDAVSSNKKRAVTV 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1530	100.0	294	22	AAV71974
2	1530	100.0	294	22	AAV71976
3	1450	94.8	281	22	AAV71975
4	576.5	37.7	376	21	AAV73340
5	324	21.2	350	21	AAV91937
6	323	21.1	350	19	AAW56264
7	323	21.1	350	21	AAV91932
8	313	20.5	350	19	AAW56265
9	313	20.5	350	21	AAV91933
10	300	19.6	286	21	AAW43527
11	300	19.6	344	21	AAW43526

12	300	19.6	372	21	AAW43525
13	299	19.5	344	19	AAW63042
14	283.5	18.5	340	20	AAW42376
15	281.5	18.4	339	16	AAW78183
16	281.5	18.4	339	19	AAW62523
17	281.5	18.4	339	21	AAW87918
18	280.5	18.3	339	15	AAW54070
19	280.5	18.3	339	16	AAW87181
20	280.5	18.3	339	19	AAW62522
21	280.5	18.3	339	21	AAW87917
22	279.5	18.3	339	22	AAW72088
23	278.5	18.2	344	22	AAW72086
24	276	18.0	365	16	AAW78182
25	271.5	17.7	340	20	AAW42375
26	243	15.9	346	21	AAW87743
27	200	13.1	189	21	AAW44428
28	153	10.0	159	21	AAW44429
29	137.5	9.0	128	21	AAW44430
30	126	8.2	781	17	AAW01871
31	113.5	7.4	461	20	AAW34463
32	113.5	7.4	462	20	AAW34339
33	107	7.0	451	13	AAW29636
34	104	6.8	386	21	AAW16691
35	103	6.7	386	20	AAW37531
36	101	6.6	273	16	AAW79145
37	101	6.6	273	19	AAW44140
38	101	6.6	347	19	AAW44138
39	97.5	6.4	198	21	AAW28750
40	97.5	6.4	384	21	AAW44386
41	96	6.3	438	18	AAW34668
42	95	6.2	763	21	AAW81751
43	95	6.2	764	21	AAW81563
44	94	6.1	487	20	AAW36888
45	93.5	6.1	340	17	AAW93198

ALIGNMENTS

RESULT 1	AAV71974	standard; Protein; 294 AA.
ID	AAV71974	standard; Protein; 294 AA.
XX	AAV71974;	
AC	AAV71974;	
XX	28-MAR-2001	(first entry)
XX	28-MAR-2001	(first entry)
XX	28-MAR-2001	(first entry)
DE	Maize RAD51 orthologue protein #1.	
XX	Maize: RAD51 orthologue; RAD52 epistasis group; double strand break; DSF;	
KW	homologous recombination; transformation; gene targeting;	
KW	transgenic plant; DNA repair.	
XX		
OS	Zea mays.	
XX	WO200068370-A2.	
PN	16-NOV-2000.	
XX	16-NOV-2000.	
PD	05-APR-2000; 2000WO-US09010.	
XX	05-APR-2000; 2000WO-US09010.	
PF	05-MAY-1999; 9905-0132582.	
XX	05-MAY-1999; 9905-0132582.	
PR	(PION-) PIONEER HI-BRED INT INC.	
XX	(PION-) PIONEER HI-BRED INT INC.	
PA	Mahajan PB, Shi J;	
XX	Mahajan PB, Shi J;	
PI	WPI; 2001-016092/02.	
XX	WPI; 2001-016092/02.	
DR	N-PSDB; AAD02118.	
XX	N-PSDB; AAD02118.	
PT	New maize RAD51 polynucleotides, useful for e.g. controlling	
PT	recombination or transformation efficiency in plants, or as probes or	
PT	amplification primers for detecting, quantifying or isolating gene	

Zea mays protein f
ATMCI1 protein seq
ZMRAD51B amino aci
Human RAD51 protel
Sequence of the sp
Human Rad51 protel
Prod. of gene part
Mouse RAD51 protel
Sequence of the sp
Mammalian Rad51 pr
Wheat RAD51 recomb
Glycine max RAD51
Fission yeast RAD5
ZMRAD51A amino aci
DNA replication an
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Recombinase enzyme
Porphyromonas ging
Porphyromonas ging
PCTD ORF 1. Chlam
Bacteriophage Dp-1
Protein Involved 1
Deduced sequence o
Helicobacter pylor
Helicobacter pylor
Arabidopsis thalia
Zea mays RecA-like
Arabidopsis thalia
Arabidopsis thalia
Streptococcus pneu
Streptococcus pneu
Protein Involved 1
Thermus aquaticus

PT transcripts -

XX Claim 11; Page 69-70; 76pp; English.

XX The present sequence is a maize RAD51 orthologue protein #1. RAD51 is one

PS of the genes of the RAD52 epistasis group that is involved in double

CC strand breaks (DSBs) repair by homologous recombination. Control of

CC homologous recombination by modulating RAD51 is used for improving

CC transformation and gene targeting in transgenic plants. The RAD51

CC nucleotide may be used as probes or amplification primers for detecting,

CC quantifying or isolating gene transcripts, in detecting deficiencies in

CC the mRNA level during screening for desired transgenic plants, for

CC detecting gene mutations, for monitoring upregulation of expression or

CC changes in enzyme activity, for detecting any number of allelic variants,

CC orthologues or paralogues of the gene, or for site directed mutagenesis

CC in eukaryotic cells. It may also be used for recombinant expression of

CC its encoded polypeptide, or for use as immunogen in preparing and/or

CC screening of antibodies, and in sense or antisense suppression of one or

CC more genes in a host cell, tissue or plant. The RAD51 proteins may be

CC used in assays to agonise or antagonise the enzyme function, or as

CC immunogens or antigens to obtain antibodies.

XX Sequence 294 AA:

SO

Query Match 100.0%; Score 1530; DB 22; Length 294;

Best Local Similarity 100.0%; Pred. No. 3.3e-157;

Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGQSSSRNGPQOKYKYSQAQNAQMFSDLSQKHITGSGDLNDLGGHCKEYETIGG 60

DB 1 mgdsgsrngpqkkyvsqagawmdsldsqkhlttgsqdlndllyggihckeyetl 60

QY 61 VPGVGKTQGLGIAINVOIPEECGLGKAVYIDTEGSMFERNVOIAEGCIRDLEHFP 120

DB 61 vpgvgktqglgialnvqipvecgglgkavayidtegsfmervyqiaegcirdllehfp 120

QY 121 HSHKSSSVQKQLQPERFLADIIYFRICSTEQIAVINMEKFLREHKDRIIIDSVPF 180

DB 121 hshksssvqkqlqperfladiyfricsteqiavinmekflrehkdriiidsvtf 180

QY 181 HFRQDFEDIALRTRVLSGLSKLTKTKTYNLAIVLLNVTTFKTEGSPQTLALGDSWS 240

DB 181 hfrqdfedialrtrvlsqslkltktylnlavllnvtctkftesqtlalagdsws 240

QY 241 HSCNRLILHMNGENERYAHLDKSPSLPVASAPYAVTGKIRDAVSSNHKRAVY 294

DB 241 hscnrlilhmngeneryahldkspslpvasapyavtgkirdavssnhkravt 294

RESULT 2

AAV71976

ID AAV71976 standard; Protein; 294 AA.

XX AAV71976;

AC

XX 28-MAR-2001 (first entry)

DT

XX Maize RAD51 orthologue protein #3.

DE

XX Maize; RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;

KW homologous recombination; transformation; gene targeting;

KW transgenic plant; DNA repair.

XX Zea mays.

OS

XX WO200068370-A2.

PN

XX 16-NOV-2000.

PD

XX 05-APR-2000; 2000MO-US09010.

PF

XX 05-MAY-1999; 99US-0132582.

PR

XX (PION-) PIONEER HI-BRED INT INC.

PA Mahajan PB, Shi J;

PI WPI; 2001-016092/02.

XX N-PSDB; RAD02120.

DR

XX New maize RAD51 polynucleotides, useful for e.g. controlling

PT recombination or transformation efficiency in plants, or as probes or

PT amplification primers for detecting, quantifying or isolating gene

PT transcripts -

XX Claim 1; Page 74-75; 76pp; English.

PS

XX The present sequence is a maize RAD51 orthologue protein #3. RAD51 is one

CC of the genes of the RAD52 epistasis group that is involved in double

CC strand breaks (DSBs) repair by homologous recombination. Control of

CC homologous recombination by modulating RAD51 is used for improving

CC transformation and gene targeting in transgenic plants. The RAD51

CC nucleotide may be used as probes or amplification primers for detecting,

CC quantifying or isolating gene transcripts, in detecting deficiencies in

CC the mRNA level during screening for desired transgenic plants, for

CC detecting gene mutations, for monitoring upregulation of expression or

CC changes in enzyme activity, for detecting any number of allelic variants,

CC orthologues or paralogues of the gene, or for site directed mutagenesis

CC in eukaryotic cells. It may also be used for recombinant expression of

CC its encoded polypeptide, or for use as immunogen in preparing and/or

CC screening of antibodies, and in sense or antisense suppression of one or

CC more genes in a host cell, tissue or plant. The RAD51 proteins may be

CC used in assays to agonise or antagonise the enzyme function, or as

CC immunogens or antigens to obtain antibodies.

XX Sequence 294 AA:

SO

Query Match 100.0%; Score 1530; DB 22; Length 294;

Best Local Similarity 100.0%; Pred. No. 3.3e-157;

Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGQSSSRNGPQOKYKYSQAQNAQMFSDLSQKHITGSGDLNDLGGHCKEYETIGG 60

DB 1 mgdsgsrngpqkkyvsqagawmdsldsqkhlttgsqdlndllyggihckeyetl 60

QY 61 VPGVGKTQGLGIAINVOIPEECGLGKAVYIDTEGSMFERNVOIAEGCIRDLEHFP 120

DB 61 vpgvgktqglgialnvqipvecgglgkavayidtegsfmervyqiaegcirdllehfp 120

QY 121 HSHKSSSVQKQLQPERFLADIIYFRICSTEQIAVINMEKFLREHKDRIIIDSVPF 180

DB 121 hshksssvqkqlqperfladiyfricsteqiavinmekflrehkdriiidsvtf 180

QY 181 HFRQDFEDIALRTRVLSGLSKLTKTKTYNLAIVLLNVTTFKTEGSPQTLALGDSWS 240

DB 181 hfrqdfedialrtrvlsqslkltktylnlavllnvtctkftesqtlalagdsws 240

QY 241 HSCNRLILHMNGENERYAHLDKSPSLPVASAPYAVTGKIRDAVSSNHKRAVY 294

DB 241 hscnrlilhmngeneryahldkspslpvasapyavtgkirdavssnhkravt 294

RESULT 3

AAV71975

ID AAV71975 standard; Protein; 281 AA.

XX AAV71975;

AC

XX 28-MAR-2001 (first entry)

DT

XX Maize RAD51 orthologue protein #2.

DE

XX Maize; RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;

KW homologous recombination; transformation; gene targeting;

KW

DB 87 alellehgqgfilifcaldldlgygypmlkteicagpygkqlcmqlavdqlp- 145
 QY 82 EC-GGLGKAVYIDTEGSPWERYOIAEGCIRD-LEHFPHSHEKSSSVOKOLOPERFL 139
 DB 146 ecfgyagvevfidegslmvdvdlataclqhlqllaeekhghehrfalefdldnll 205
 QY 140 ADIYFRICSYTEQIAVINMEKLEHKKVDRIIVDSVTFHFRDOFEALARTVLSGL 199
 DB 206 shlyycticdytelqlavylldpflshskvrlvldqldafpfrldldsltrllngl 265
 QY 200 LSKLMKIAKTYNLAAYVILNQVTFTEGSPQTLALGDSWSHSCNRLILHNGNERVAH 259
 DB 266 agqmislannhrhavlhtgmktkldngallvpalgewgnhaatrlrlfhwdrkqrilat 325
 QY 260 LDKSPSLPVASAPYAVTGKIRDAV-----SSNHRKAR 292
 DB 326 lykspqkectvlfqikpqgfrdvtvsacslqtqegslstrkrxr 370

RESULT 5
 ID AAY91937 standard; Protein; 350 AA.
 AC AAY91937;
 DT 19-JUL-2000 (first entry)
 DE Ala163 modified human Rec2 serine kinase.
 KW hsrRec2; serine kinase; homologous pairing; strand transfer; RAD51;
 KW phosphorylation; cyclin E; p53; cell cycle.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc-difference 163 /note= "Y163A substitution from wild type"
 FT MO200017329-A1.
 PN 30-MAR-2000.
 PD 17-SEP-1999; 99WO-US21642.
 PF 21-SEP-1998; 98US-0157603.
 PR (UYJR-) UNIV JEFFERSON THOMAS.
 PA (CORR) CORNELL RES FOUND INC.
 PA (KIME-) KIMERAGEN INC.
 PI Havre PA, Rice MC, Holloman WK, Kmiec EB;
 DR MPI; 2000-283562/24.
 PT phosphorylating a serine-containing substrate by incubating it with
 PT adenosine triphosphate and Rec2 kinase and measuring the level of
 PT phosphorylation, useful for discovering specific antagonists or
 PT agonists of Rec2
 PS Claim 18; Page -; 41pp; English.

CC The present sequence is Ala163 hsrRec2, a human serine kinase modified by
 CC a Y163A substitution. The hsrRec2 is in the same supergene family as the
 CC mammalian protein having homologous pairing and strand transfer
 CC activities, RAD51 and was isolated because of its homology to the
 CC homologous pairing and strand transfer protein of *Ustilago maydis*. In
 CC particular, hsrRec2 phosphorylates several proteins that control the cell
 CC cycle, especially cyclin E and p53. The invention permits the
 CC phosphorylation of cell cycle control proteins at sites that are
 CC physiologically relevant. The invention can be practiced with either
 CC murine or human Rec2 or a mutin or chimera of these proteins. In
 CC particular the mutin has the sequence of a Rec2 kinase containing other

CC than a Tyr at position 163. The invention comprises a method of
 CC phosphorylating a serine-containing substrate comprising incubating the
 CC substrate with ATP (adenosine triphosphate) and Rec2 kinase or a
 CC mammalian Rec2 and measuring the level of phosphorylation. The method is
 CC useful for discovering compounds which are specific antagonists or
 CC agonists of Rec2.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from the hsrRec2 sequence given on pages 33-34
 CC
 SO Sequence 350 AA;

Query Match 21.2%; Score 324; DB 21; Length 350;
 Best Local Similarity 30.9%; Pred. No. 1.4e-26;
 Matches 88; Conservative 50; Mismatches 99; Indels 48; Gaps 7;

QY 27 SDELSOKHITGSGDNDLIDGGHCKEVTGCGVPGVGTQLGQALAINVOLPVECGGL 86
 DB 75 sadfspalstlslaldealngvacgslteltpgpcgkqfclmmlslatlpnmgl 134
 QY 87 GSKAVYIDTEGSPWERYOIAEGCIRDLEHFPHSHEKSSSVOKOLOPERFL---ADIV 143
 DB 135 egavvyldtesaefaerlveiae-----srfrafnte-----ekllltskvh 178
 QY 144 YFRICSYTEQIAVINMEKLEHKKVDRIIVDSVTFHFRDOFE----DLALTRVLSG 198
 DB 179 lyreltdevlqrlsleeeis-kglklvldsvavsvkefdagqlgnlkerkfilar 237
 QY 199 LSKLMKIAKTYNLAAYVILNQVTFK-----TEGSPQTLALG 236
 DB 238 easslkylaeefslpvltnqtlhslgalasqadlvsppaddslssegtsysscvcvlaig 297
 QY 237 DSWSHSCNRLILHNGNERVAHL-DKSPSLPVASAPYAVTGKCI 280
 DB 298 ntwshtvntlllqyldserqlllaksplapftsftytlkegl 342

RESULT 6
 ID AAM56264 standard; Protein; 350 AA.
 AC AAM56264;
 DT 28-AUG-1998 (first entry)
 DE Human recombinase hsrRec2 protein.
 KW Recombinase; REC2; hsrRec2; homologous recombination; irradiation;
 KW apoptosis; cancer; human.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 4..7 /label= NLS
 FT Region /note= "nuclear localisation sequence"
 FT Region 108..116 /label= A-box
 FT Region /note= "A box motif sequence"
 FT Region 160..163 /note= "Src-type phosphorylation site"
 FT Region 186..199 /label= DNA
 FT Region /note= "DNA binding sequence"
 FT Region 206..210 /label= B-Box
 FT Region /note= "B box motif"

PN W09811214-A1.
 PD 19-MAR-1998.
 PF 11-SEP-1997; 97WO-IB01217.

CC mammalian Rec2 and measuring the level of phosphorylation. The method is
CC useful for discovering compounds which are specific antagonists or
CC agonists of Rec2.

SQ **Sequence** **350 AA;**

Query Match	20.5%;	Score 313;	DB 21;	Length 350;
Best Local Similarity	32.1%;	Pred. No. 2.2e-25;		
Matches 100; Conservative	46;	Mismatches 98;	Indels 68;	Gaps 13

```

QY 20 QANMWD----FSELSOKHITNGSSDLNDIDGGHCKEYTBEGVPGVKQIOLAIIN 76
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 65 qtaueiktrrshispaistlftlcaldeaahgvyrcgssteltpprgcykqcficmmsvl 124
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 77 WQIPVECGGLGKAVYIDTEGSEFWERVYOAIEGCIJDILEHP--HSHER---SSSV 139
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 125 atlpalaglegavuytltesaftaelvaeia-----srfpyfntkeallltsarv 177
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 130 Q--KQAPRFLADIVYRIGSYTEQOIAVINWEKFLREHKVRIVIIDSYFPHRODFE 187
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 178 hlcrlctcgllq-----lksleeeell-----skvklvldgsiasvvrtefd 221
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 188 -----DLALRTVYL-SGSLKLMKIATKYALVALLNQVTRF-----224
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 222 pklqgnlkerunkflfgkasl-lkylaeafispyllnqtlhtlsagdlpsqadlvsppddl 280
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 225 -----TEGFSQTLTALGSMGSHSCTNNLIMHMGNGERYAHL-DKSPSLPVASAPYAVTQK 278
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 281 slseigsagsscaalagntwchvntcllllqyldserqjllakspriaafscfyytlkge 340
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 279 GIRDVASSHKKR 290
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 341 gl---vlqgher 349
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 10

ID	AAG43527	standard; Protein; 286 AA.
----	----------	----------------------------

AC MAG43527;

DT	18-OCT-2000	(first entry)
XY		

Zea mays protein fragment SEQ ID NO: 54415

KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter
KW termination sequence; corn.

Zea mays subsp. *ma*

PN EP1033405-A2

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439

PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.

PR	30-APR-1999	9905-0132048
PR	30-APR-1999	9905-0132047
PR	04-MAY-1999	9905-0132484
PR	04-MAY-1999	9905-0132485
PR	05-MAY-1999	9905-0132486
PR	06-MAY-1999	9905-0132487
PR	07-MAY-1999	9905-0132488
PR	11-MAY-1999	9905-0134256
PR	14-MAY-1999	9905-0134218
PR	14-MAY-1999	9905-0134219
PR	14-MAY-1999	9905-0134221
PR	14-MAY-1999	9905-0134370
PR	18-MAY-1999	9905-0134761
PR	19-MAY-1999	9905-0135124
PR	20-MAY-1999	9905-0135125
PR	21-MAY-1999	9905-0135553
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136392
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	10-JUN-1999	9905-0139119
PR	14-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139465
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0139617
PR	22-JUN-1999	9905-0139699
PR	23-JUN-1999	9905-0140353
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140623
PR	24-JUN-1999	9905-0140625
PR	29-JUN-1999	9905-0140891
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	01-JUL-1999	9905-0142154
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0144220
PR	12-JUL-1999	9905-0144327
PR	13-JUL-1999	9905-0144352
PR	14-JUL-1999	9905-0144362
PR	15-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	16-JUL-1999	9905-0144225
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144333
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144632
PR	20-JUL-1999	9905-0144884
PR	21-JUL-1999	9905-0144814

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132047.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139455.
PR 16-JUN-1999; 99US-0139455.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139765.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154079.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 19.6%; Score 300; DB 21; Length 344;
Best Local Similarity 33.1%; Pred. No. 5,5e-24;

Matches 88; Conservative 41; Mismatches 109; Indels 28; Gaps 5;

OY 24 DMFSELISQKHITGSGDLNDILGGCHKEVTEIGVGVGKQGLQALINQVPEC 83
DB 97 d1llkrksvntltgsgalndelggltetlciteafgfsrgtqlahltcvtqlp1hm 156
OY 84 GGLGKAVYIDPESSEWERYOIAEGCIRDLLEHFRSHKSSVQKOLQPEFLADY 143
DB 157 hgngskvayidtegtffperlvplae-----ffgmnaanavlhl1 196
OY 144 YERICSYTEQIAYINMYMEKFLREHKDVRIVYIDSVTFHRODFE---DLALRFTVLSGLS 200
DB 197 Yaretyehgynlljglaakmae-epfklllvdsvalfrvdfsgrelaerqk1agml 255
OY 201 LKMKIKAKTYLAVVLLNQVTKFTBESF---QUTLALGDSWSHSCNRLILH-WNGNER 256
DB 256 srltklaeeftvayltngv1adp9gsmfildpkkpagnh1aactlrlmlrkykgegr 315
OY 257 YAHLDKSPSLPVASAPYAVTGKGRD 282
DB 316 vcklfadpnlpegeavfqtvsagimnd 341

RESULT 12

ID AAG43525 standard; Protein; 372 AA.

XX AAG43525;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 54413.

KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence; corn.

XX Zea mays subsp. mays.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127452.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.


```
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
```

```
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156599.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 19.6%; Score 300; DB 21; Length 372;
Best Local Similarity 33.1%; Pred No. 6.1e-24;
Matches 88; Conservative 41; Mismatches 109; Indels 28; Gaps 5;

```
QY 24 DMFSDLSQKHITTSQSGNDILGGICRCKEYETIGVPGVKTOLGQLAINVQIPVC 83
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 125 dlkkksvrlttsqalldellggilecltfeafgrstqtlahlcstqlplhm 184
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 84 GGLGKAVIITDEGSPWVRVYQIAGCTRIDLEHPPSHKSSSVQKOLQPERLADY 143
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 185 hgnngkvayidtegltrpivplae-----rfgmnanavldhll 224
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 144 YFRICSTEOIAVINMEKFLRHKDVRIVIIDSVTFHRDPE--DLALRTVLSGLS 200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 225 yaravlyehqynlllglaakmae-epfkllivdsvaltrvfsgrelaeqklaqm 283
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 201 LKIMIAKTYNLAVLNLOVTTKFTGSGF---OLTLALGDSMSHCTNRLILH-MNGNER 256
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 284 srltklaeefnvayltlnqviadppgsmfildpkkpaggvnlahaatitlmrlkkggejr 343
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 257 YAHLDKSPSLPVASAPYATVGKGIKD 282
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 344 vcklfadnlpgeavfyfytsgimnd 369
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 13
AAM63042
ID AAM63042 standard; Protein; 344 AA.
XX
AC AAM63042;
XX
DT 06-OCT-1998 (first entry)
XX
DE ATDMCI protein sequence.
XX

Tue Oct 2 11:16:27 2001

us-09-537-654-2.rag

Page 14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 04:52:04 ; Search time 2002.11 Seconds
(without alignments)
11387.717 Million cell updates/sec

Title: US-09-537-654-1
Perfect score: 1474
Sequence: 1 tcgaccacgcgcgcgcact.....aaaaaaaaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_da1:*
5: gb_da2:*
6: gb_da3:*
7: gb_da4:*
8: gb_da5:*
9: gb_da6:*
10: gb_da7:*
11: gb_da8:*
12: gb_da9:*
13: gb_da10:*
14: gb_da11:*
15: gb_da12:*
16: gb_da13:*
17: gb_da14:*
18: gb_da15:*
19: gb_da16:*
20: gb_da17:*
21: gb_da18:*
22: gb_da19:*
23: gb_da20:*
24: gb_da21:*
25: gb_da22:*
26: gb_da23:*
27: gb_da24:*
28: gb_da25:*
29: gb_da26:*
30: gb_da27:*
31: gb_da28:*
32: gb_da29:*
33: gb_da30:*
34: gb_da31:*
35: gb_da32:*
36: gb_da33:*
37: gb_da34:*
38: gb_da35:*
39: gb_da36:*
40: gb_da37:*
41: gb_da38:*
42: gb_da39:*
43: gb_da40:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_v13:*
61: gb_v14:*
62: gb_v15:*
63: gb_v16:*
64: gb_v17:*
65: gb_v18:*
66: gb_v19:*
67: gb_v20:*
68: gb_v21:*
69: gb_v22:*
70: gb_v23:*
71: gb_v24:*
72: gb_v25:*
73: gb_v26:*
74: gb_v27:*
75: gb_v28:*
76: gb_v29:*
77: gb_v30:*
78: gb_v31:*
79: gb_v32:*
80: gb_v33:*
81: gb_v34:*
82: gb_v35:*
83: gb_v36:*
84: gb_v37:*
85: gb_v38:*
86: gb_v39:*
87: gb_v40:*
88: gb_v41:*
89: gb_v42:*
90: gb_v43:*
91: gb_v44:*
92: gb_v45:*
93: gb_v46:*
94: gb_v47:*
95: gb_v48:*
96: gb_v49:*
97: gb_v50:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1474	100.0	1474	9	AX046601 Sequence
2	1305.4	88.6	1333	9	AX046605 Sequence
3	1243.2	84.3	1459	9	AX046603 Sequence
4	358.8	24.3	1101	13	ATN299424
5	178.4	12.1	1295	88	AF029669
6	171	11.6	13652	83	AP003310
7	171	11.6	163419	13	AP002908
8	103.4	7.0	1123	91	BC000667

9	101.4	6.9	2340	13	ATH296174	AJ296174 Arabidops
10	101.4	6.9	122871	12	AC002387	AC002387 Arabidops
11	87.2	5.9	110000	84	LMFCH32_17	Continuation (18 o
12	77.8	5.3	10980	1	AE001035	AE001035 Archaeogl
13	74.6	5.1	13968	1	AE000901	AE000901 Methanoba
14	72	4.9	1699	3	SS045310	U43310 Sulfolobus
15	71.4	4.8	202000	2	AP000058	AP000058 Aeropyrum
16	70	4.7	1241	14	LEU22441	U22441 Lycopersico
17	68.6	4.7	1995	14	CAU39808	U39808 Candida alb
18	67.2	4.6	1740	14	DE4035	DE4035 Schizosacch
19	67.2	4.6	4230	12	AB008545	AB008545 Schizosac
20	67.2	4.6	15350	11	SPAC8E11	AL021817 S. pombe c
21	66.4	4.5	844	71	AC029940	AC029940 Giardia 1
22	66.4	4.5	992	75	AC074448	AC074448 Giardia 1
23	66.4	4.5	1030	72	AC058893	AC058893 Giardia 1
24	66.4	4.5	1050	75	AC078679	AC078679 Giardia 1
25	65.8	4.5	1551	12	AF079428	AF079428 Zea mays
26	65	4.4	1585	9	AX047126	AX047126 Sequence
27	64.4	4.4	1551	12	AF079429	AF079429 Zea mays
28	64.2	4.4	813	53	CNS075T2	AF198107 Giardia 1
29	63.6	4.3	2745	5	AF198107	AF198107 Giardia 1
30	62.8	4.3	1177	14	ATU43528	U43528 Arabidopsis
31	62.8	4.3	1219	9	AX047122	AX047122 Sequence
32	62.6	4.2	1800	15	YSCDMC1A	AF017729 Oryctolag
33	62.4	4.2	2774	15	YSCDMC1A	AF017729 Oryctolag
34	62.4	4.2	2774	15	YSCDMC1A	AF017729 Oryctolag
35	62.4	4.2	64044	14	SCE9163	DI0865 Yeast genes
36	61.2	4.2	1518	1	AB026115	AB026115 Pyrobacul
37	60.6	4.1	921	72	AC066858	AC066858 Giardia 1
38	60.6	4.1	928	72	AC045789	AC045789 Giardia 1
39	60.2	4.1	1101	72	AC053814	AC053814 Giardia 1
40	59.8	4.1	1260	72	AF145465	AF145465 Desulfuro
41	59.4	4.0	1502	1	AF052597	AF052597 Pyrococcu
42	58.6	4.0	923	53	CNS060D1	AL407915 T3 end of
43	58.4	4.0	1021	53	CNS060D1	AL411887 T7 end of
44	58.4	4.0	1021	53	CNS060D1	AL410286 T3 end of
45	58.4	4.0	147620	62	AC011195	AC011195 Homo sapi

ALIGNMENTS

RESULT	1	1474 bp	DNA	PAT	15-DEC-2000
AX046601	Sequence 1	from Patent WO0068370.			
LOCUS	AX046601				
DEFINITION	AX046601				
ACCESSION	AX046601				
VERSION	AX046601.1	GI:11876173			
KEYWORDS					
SOURCE					
ORGANISM					

Ze mays.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 1474).
 Mahajan, P.B. and Shi, J.
 Rad51 from maize
 Patent: WO 0068370-A 1 16-NOV-2000;
 PIONEER HI-BRED INTERNATIONAL, INC. (US)
 Location/Qualifiers

FEATURES

source

1..1474

/organism="Zea mays"

/db_xref="taxon:4577"

310..1194

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC18901.1"

/db_xref="GI:11876174"

/translation="MGDSGRNGPQKRYSGAQNAMDFSDLSQKHITGSGDLND
 ILGGICKEVTEIGGVGKGTGLGIAIIVQIPVCGGSGKAVYIDEGSPME
 RYVQIAEGCIRIDLEHPSHEKSSVQKLOPEFLADYIFRCSTYEDIAVNYM
 EKPIRHKVRIVITDSVTFHFRDOPEDIALPTRYLSGIKIMKIACTYINAVILN
 QVTTFTBESFQITLALGDSMSHCTNRLILHMNGNERAHLDKPSLPVASAPAVT

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY 961 aaccaagtcactactaaattacagaaggtcattcattgaactctgtctagtgac 1020
|||||
Db 961 AACCAAGTCACTACTAAATTACAGAGGCTCATTTGACTCTTGCTCTAGTGAC 1020
QY 1021 agctggtccactatgacagaaccggttattctgcactcgtgaatggaacgaatc 1080
|||||
Db 1021 AGCTGCTCCACTATGACAGAACCGGTGATTTCTGACATGGAATGGGAACGATAC 1080
QY 1081 gcaactctgaatgctccctcactccactgaagcccaagcagatgagtcagcagc 1140
|||||
Db 1081 GCACATCTTGATTAAGTCTCCCTTCACTTCCAGTACAGCAGCCGATGACAGTAC 1140
QY 1141 aaaggatagaagatgctgtagctcaaacacagaagcgaagccgagtaacgattct 1200
|||||
Db 1141 AAAGGATTAAGATGCTGTAGCTCAACACAGGAGAGCCGATTAAGTACATTC 1200
QY 1201 tgggtcaagacttgatgtccactacgtctcctcgaagcttctgcacatgacatctt 1260
|||||
Db 1201 TGGTCAAGCACTTGTATGCTCACTACGCTTCCGACGCTTCTCCGATGATCTTT 1260
QY 1261 ggaactgaagtgagactgagagatagtaaccatttctgtagtctcagttgtctg 1320
|||||
Db 1261 GGACTAGAGAGTGAAGTGAAGATAGTACATTTGTTGATTTCTAGTGTCTTGTGC 1320
QY 1321 cgttggtaccacacacacttaagagaagtaataacacagaagcgttaatatgtg 1380
|||||
Db 1321 CGTTGGCTACCAACCACTTAAGAGAGATTAATACACAGAACGAGCTAATATAGTG 1380
QY 1381 ttgtgtatgaactgtgcccactgtaactatagtaagccatataatgagcgcaaaa 1440
|||||
Db 1381 TTTGTATCTGAACATCTGCGCCATGCTACATTCATGAAGCCATATATAGCGGCAAAA 1440
QY 1441 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474
|||||
Db 1441 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1474

RESULT 2
AX046605 1333 bp DNA PAT 15-DEC-2000
LOCUS Sequence 5 from Patent WO0068370.
DEFINITION AX046605
ACCESSION AX046605
VERSION AX046605.1 GI:11876177
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1333)
AUTHORS Mahajan, P.B. and Shi, J.
TITLE Rad51 from maize
JOURNAL Patent: WO 0068370-A 5 16-NOV-2000;
PIONEER HI-BRED INTERNATIONAL, INC. (US)
FEATURES
source
1.1333
Location/Qualifiers
/organism="Zea mays"
/db_xref="taxon:4577"
169..1053
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC18903.1"
/db_xref="GI:11876178"
translation="MGDQSGSRNGPOKXVSGNANWDFSDLSOKHTTGGSLND
ITGGTHCKEVEITGSGVGVKTOGLAIWQIPVECGGAGKAVYIDTGSFWE
RYQIAEGCTRIOLIEFPHSHKSSSVQKQLOPERLADITFRICSTEQIAYINM
EKFREHKDVRIVITDSVTFHRQDEIDALTRRVLSGSLIMKIAKYNLAVALN
OYTKPTBGSFQITLIDGWSHSCNRLILHMNGERNERAHDKSPSLPVASAPAVT
GKIRDAVSNNHRAVY"
BASE COUNT 377 a 284 c 326 g 345 t 1 others
ORIGIN

Query Match 88.6%; Score 1305.4; DB 9; Length 1333;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1315; Conservative 1; Mismatches 17; Indels 0; Gaps 0;
QY 142 cgaagcgaagcgcgtgcgtgagccacgacgagcgtacgagcgtgtgtgagcagata 201
|||||
Db 1 CGACGTAAAGGGGTGCGTGGCGCCACGAGGAGGTACGACGGTGTGGAGCAGATA 60
QY 202 ttggaagtggaagtgctcaaacggttcggcggtgtgagatctgaatccgcactga 261
|||||
Db 61 TGAGAGGTGAGGTGCTCAACAGGGGTGCGGCTGTGATCTGAATTCGCACTGTA 120
QY 262 gtctctcttcccccacatcagtaaccctcccaagtgagcaatcccaatggagataca 321
|||||
Db 121 GTTCTCTTCTTCCCAATACAGTACACCTTCGCAAGTGGCAATGACATGGAGATCA 180
QY 322 tctgctcttagaaatgagccacaacagaagtaagttcagaagcccaagaatgcctggat 381
|||||
Db 181 TCTGGCTCTAGAAATGACACACAGAAATGATTCAGGAGCCGAGATCCTGGGAT 240
QY 382 atgtctctgagagtgctcagagaacacatcactagtgctcgtgtaacctcaatgac 441
|||||
Db 241 ATGTTCTTGATGAGCTGTGACAGAAACATCACTACTGTTCTGTGATCTCAATGAC 300
QY 442 atacttggtgcygagatcactgcaagaagatctagatcgtgcygtcccaagggtt 501
|||||
Db 301 ATACTGTGTGGCGGATTCATGCAAGAAAGATTACTGATATCGGTGGCGTCCAGGGGT 360
QY 502 gtttaaacctcaactggtggtatcaactagcaatcaatgtcaaatcccaagtgatgtgt 561
|||||
Db 361 GGTAAACCACTCACTGGGATTCACATGACCAATCAATGCAATCCCATGGAATGTGT 420
QY 562 ggccttggtggaagcaattcatatagatagagagggagttcaatgtgtgaagctgc 621
|||||
Db 421 GGCTTGGTGGGGAACCACTTTATATAGTATACAGAGGAGCTTCCGACAGCATGAG 480
QY 622 taccagatctgtaaggtgtgataataggaacactggaagcattccgcacagacatgag 681
|||||
Db 481 TACCAATGTTGTAAGAGGGGTATAGGACATATGAGACACTTCCGACAGCATGAG 540
QY 682 aagtcctctctgccaacaaatatacagcgtgagcgtttcctgtgagatatactaac 741
|||||
Db 541 AAGTCTCTTCTGTCGCAAAACAAATTCAGCTGAGCGCTTCTGCGGATATCTATTAC 600
QY 742 ttccgatatgcaatgacacgaacaaatgcaatgcaatgcaatgcaatgcaatgcaat 801
|||||
Db 601 TTCCGATATGCAAGTATACCCGACAAATTCGATCATTAATGAGAGATTTCTC 660
QY 802 agagagcataaagatgtgcgtatagttatattatgataagtgtaacttccacttcgacaa 861
|||||
Db 661 AGAGACATTAAGATGTGGGTATAGTTATTTGATAGTGTACTTCCACTTTCGACAA 720
QY 862 gatttgaagatctgagcactgagagcagagatgctcaagtgtgatatcatgaaatg 921
|||||
Db 721 GATTTTGAAGATCTGCGACTGAGACCAAGATGCTTAAGGATTTATCTTAAGTTAATG 780
QY 922 aagattgcaagacatatataacttggcagttgtctgttgaacaaagtaactactaattt 981
|||||
Db 781 AAGATTGCAAAACATATATACCTTGGCAGTGTCTGTGTAACCAAGTCACTTAATTT 840
QY 982 acagaaggtcattcaattgactctgtcttaggttgacagctgtgtccactactgacag 1041
|||||
Db 841 ACAGAAAGGTCAATTCATTAATGACTCTGTGCTAGTGACAGCTGTCACATGACAG 900
QY 1042 aacgggttgatctgcaactggaatggaagacagatacgcacatctgataagctctct 1101
|||||
Db 901 AACCGTTCATCTCCACATGGAATGGAACGAACATACGACACATCTGATTAAGTCTCT 960
QY 1102 tcaacttcaagtagctcagcaccgctatcgaatggaagcaaggaataagagatctgtg 1161
|||||
Db 961 TCACTTCCAGTAGCCCTACACACCGTATGACAGGCAAGGAGATTAAGATCTCTGTG 1020
QY 1162 agtcaaacacacagcagccgagtaacgtatctctgtgtgcaagcactgtatgt 1221

```

Db 1021 AGCTCAACACACAGGACCCGAGTAAGTACATTCCTGGTGTGCAAGCAATTTATGT 1080
Qy 1222 ccaactagctcctgcagctcttcttcgcatagtatcttlttgactagttagagtgag 1281
Db 1081 CCACCTACGCTCCGCTGCTCTTCTGCCCAGTATGATCTTTGGAGTGTAGGAGACCTGG 1140
Qy 1282 agaatgataccatttggatcttcagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1341
Db 1141 AGAATGATACATTTGTTGATTTTCAGTTGCTTGTGCGCTTGGCTACCAACCACTT 1200
Qy 1342 aagagagaagtaataacacagagagctataatagtggttlttgatactgaacatctggc 1401
Db 1201 AAGAGAGAGTAATTAATCAACAGACAGGCTAATATAGTGTGTTGTTGATCGAACATCTGGS 1260
Qy 1402 ccatctacattcagtaagaagcctataatagcggcgaataaaaaaataaaaaaataaaa 1461
Db 1261 CCATCTACATTCAGTAAGCCATTAATATAGGCGCATATATGTGCTTCTCTGATCAAAA 1320
Qy 1462 aaaaaaaaaa 1474
Db 1321 AAAAAAAAAAAAAA 1333

RESULT 3
LOCUS AX046603 1459 bp DNA PAT 15-DEC-2000
DEFINITION Sequence 3 from Patent WO0068370.
ACCESSION AX046603
VERSION AX046603.1 GI:11876175
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; PanC clade;
Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1459)
AUTHORS MahaJan,P.B. and Shi,J.
TITLE Rad51 from maize
JOURNAL Patent: WO 0068370-A 3 16-NOV-2000;
PIONEER HI-BRED INTERNATIONAL, INC. (US)

FEATURES
source
1..1459
/organism="Zea mays"
/db_xref="taxon:4577"
169..1014
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC18902.1"
/db_xref="GI:11876176"
/translation="MGDQSGRNGPQKRVVSGAONAMDMFSDLSOKHITTTGSDLND
ILGGIHCKEVTEIGVGAVGKTOGLAINVOIPVCGGLGKAVYIEGSPWERY
VOIAREGTRIDIEHPHSHKSSVOKOLOPERFLADYIFRISYTOIIVINMEK
PIREHKDVRIVITDSVPHFRODPEDIALRTIRVLSGLTKMKIAKTNTLAIVLLNOY
TTKPTGESFQLTLALGDSWSHSCTRLLILHNGNRRYAHLDKPSLPAVASPYATVK
GTRDY"
BASE COUNT 504 a 288 c 326 g 341 t
ORIGIN
Query Match 84.3%; Score 1243.2; DB 9; Length 1459;
Best Local Similarity 98.9%; Pred. No. 8.1e-290;
Matches 1286; Conservative 0; Mismatches 3; Indels 11; Gaps 3;

```

```

Db 121 GTTCTCTTCCCAATGATGATGACACCTCTCCAAAGTGCAATGACATGGAGATCAA 180
Qy 322 tctgtccttagaataatggaacacacagaagaatgaagttcgaagacccaagaatgtcctggat 381
Db 181 TCTGTGCTCTAATAATGGAACCAACACAGATGATGCTTTCAGAGCCCAAGAAATGCTGGAT 240
Qy 382 atgtctctgtatgagcgtctacagaaacacatacactactgtgtctgtgtgaacctaaatgac 441
Db 241 ATGTCTCTGTATGAGCGTGTACACAAACATCATCTACTGCTGTTCTGGTGTACATGAC 300
Qy 442 atacttggtggtggtggttctcactgcaaaagaattactgagatcggtgtgtgtgtgtgtgt 501
Db 301 ATACTTGTGTGGCGGATTCACCTGCAAGAAAGTACTGATGATCGGTGCGCTCCAGGGGT 360
Qy 502 gataaactcaactggtggtggttcaactagcaatcaatgtacaatcccaagtggaatgtgt 561
Db 361 GGTAAACTCACTGGGATTCACACTGACATCATATGACAAATCCAGTGAATGTGT 420
Qy 562 ggccttggtggaagagattatataatagatacagaaggttcatatgtgtgaagtgatc 621
Db 421 GGCTTGTGGTGAAGAGCATTTATAT-----AGAGGCGAGTTTCATGTTGAACGTGTC 474
Qy 622 taccagattctgaaggtgtatataaggaacatactggaagcaattccgcacagccatgag 681
Db 475 TACCAGATTGCTGAAGGCTGATTAAGGACATATCTGAGCAGCTTCCGACACACCATGAG 534
Qy 682 aagtcctctctgtccaaaacaaatcacgcctgaggtcttccctggtggtatcctatc 741
Db 535 AAGTCTCTTGTGCAAAAACAAATTAACGCTGAGCGTTTCCCTGGGGATATCTATTAC 594
Qy 742 ttccgatalgcaatcacccgaaacaaatgcagtlcaaaactacatgagaatctctc 801
Db 595 TTCCGATATGACATGATACACGCAAAATGACATGATCAATCAATGAGGAATGCTTC 654
Qy 802 agagaacataaagatgtgcgtatagttatagatagtggttctccttccacttgacaa 861
Db 655 AGAGAGCATTAAGATGTCGATATGATTAATGATGATGATGATGATGATGATGATGATGAT 714
Qy 862 gatttgaagatctgcagactgagacagagtgctgaagtgatlatcatgaaatgaatg 921
Db 715 GATTTGAAGATCTGGCAGCTGAGCAGCAGAGTGTGATGATGATGATGATGATGATGATGAT 774
Qy 922 aagattgcaagacataatactgtgcagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 981
Db 775 AAGATTGCAAGACATTAATTAATGCAATGCTGCTTGTGAACCAAGCATCTACTTAATTT 834
Qy 982 aagaaaggtcatctcaattcaattgactctgtccttagtgtaacagctgtgtccactatgacg 1041
Db 835 ACAGAGGCTCATTTCAATTGACTCTTGTCTTAGGTACAGCTGTGCTCCACTCATGTCAGC 894
Qy 1042 aacggttgcattctgcactggaatggaagaaacagatacagacatctgtaaaagctcct 1101
Db 895 AACGGTGTGATTCGACATGGAATGGAGAACAGATACGACATCTTTGATTAAGTCTCT 954
Qy 1102 tcaacttcagtaagcctgaagcagcgtaigcagtgacagcaaggaatgaagatgtcgtg 1161
Db 955 TCACCTCCAGTAGGCTCAGCCCGCTATGACAGTACAGCAAGAGGATTAAGATG- TGTG 1013
Qy 1162 agctcaaaccaacagagagcccgagtaacgtatgactctgtgtgtcaagcaactgtatgt 1221
Db 1014 AGCTCAAAACCAAGCGAGCCGAGTAAGCATGATCTTGTGTCAAGCACTTGTATGT 1073
Qy 1222 ccaactagctcctgcagctcttcttcgcatagtatcttlttgactagttagagtgagactgg 1281
Db 1074 COACTACGCTCTGAGCTTCTTCTGCGCATGATCTTTGGACTAGAGAGTGTGAGACTGG 1133
Qy 1282 agaatgataccatttggatcttcagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1341
Db 1134 AGAATGATACAT-----TTGATCTCAATGCTTGTGCGGTGGCTACCAACCAACTT 1189
Qy 1342 aagagaagaataataacagaaacagcctaataatagtggttlttgatactgaacatctggc 1401

```


Db 1190 AAGAGAGTAATACACAGAGGCTAATATGTTTGTATCTGACATCTGGC 1249

QY 1402 ccatcgtacatcagtaaacgctataatagcgaggaanaa 1441
|||||
Db 1250 CCATGCTACATTCAGTAAAGCCTAATATAGCGGCATATA 1289

RESULT 4
ATH299424
LOCUS Arabidopsis thaliana mRNA for RAD51C protein. 20-OCT-2000
ACCESSION AJ299424
VERSION AJ299424.1 GI:10944744
KEYWORDS RAD51C gene.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1101)
AUTHORS Slaud, N.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1101)
AUTHORS Slaud, N.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2000) Slaud N., Reparatoin/Recombinaison de l'ADN chez Arabidopsis, Institut de Biotechnologie des Plantes, Batiment 630, Universite Paris XI 91405 Orsay Cedex, FRANCE
LOCATION/Qualifiers

FEATURES
source 1..1101
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
gene 1..1092
/chromosome="II"
CDS 1..1092
/gene="RAD51C"
/gene="RAD51C"
/function="recombination factor"
/codon_start=1
/evidence="experimental"
/product="RAD51C protein"
/protein_id="CAC14091.1"
/db_xref="GI:10944745"
translation="MISFGRKSPRIETSLASVYAMPLPISIRKLIAGYTC
LSTIASVSSDLADANITEAEFILLKLNOSCSNRLINGAKNAMPHEESL
PRITSCSDLDNIGGISCQDVTEIGVPGIGTQIGLQSVNVDIPREGSLGGA
IYIDTESFVWKALQIAEACVEDMEYEMHHRFOANOVOMKPEDILENIFPRVC
SYEQIALVNHLEKFIENKDVKYIVDSITFHRQDPLAORTRYLSMALKFMKL
AKKSLAVVLLNVTTFKESGSPOLATALGDSMSHSCITNRIILYMNDEKRAYIDKSP
SLPSASVYTVTSRGLNSSSSSKRVKM"

BASE COUNT 307 a 219 c 268 g 307 t
ORIGIN

Query Match 24.3%; Score 358.8; DB 13; Length 1101;
Best Local Similarity 65.7%; Preg. No. 3.8e-76;
Matches 522; Conservative 0; Mismatches 272; Indels 0; Gaps 0;

QY 361 ggaagccagaaatgctgagatgtctctctgtagagctgtcaagaacacatcactact 420
|||||
Db 262 GGAGCAAGAACGCTGGGATGCTTCACAGAGAGAGTCTTCCCGATATCTACACA 321

QY 421 ggtctgtgagactaaagacatctgtgtgagagatcactgtaagaagatctactgag 480
|||||
Db 322 TCTTGTCTGATCTTGATACATTTTGGCGGTGAATTAAGCTGTAAGGATTTACAGAG 381

QY 481 atcgtgtgctccagggggtgtgtaaacacacggtggtatcactgacatcaatgta 540
|||||
Db 382 ATTGGTGGGTACACAGGAGTTGGCAAGACTCAGATTGGGATCCACCTCTCTGTAATGTT 441

QY 541 caatccagtgagatgtgtgctgtgtggaagacagattatataatgataagaagc 600
|||||

Db 442 CAGATTCACGTCGAGTGTGTGTGTGGAGGGAAGCTATATATATGATACAGAGGT 501

QY 601 agltcatgtgtgaacgtgtctaccagatgtctgaagggtatataaggacatctagag 660
|||||
Db 502 AGCTTCATGTGTGAAGGCTCTTACAGATAGCAGAGGCTTGTGTAGAGACATGGAAGA 561

QY 661 cacttccgacagagcagatgagagctctctctctgtccaaaacaaatlaagcctgagct 720
|||||
Db 562 TACACAGATACATGTCATTAACATTTTCAAGCAAAATCAAGTACAAATGAAACACAGAGT 621

QY 721 ttcctgagatatactatctactctccgatatgcaatgacacccaacaaatgacatcata 780
|||||
Db 622 ATCTTGAAGACATATTTACTTCCGTCTGTCAGTTTACACCGACGCAATCCGATTTGGTC 681

QY 781 aactatgagagagctctctcagagagcataagatgtgctatagttatctatagat 840
|||||
Db 682 AATCATCTTGAAAGATTCATCTGTGAAACCAAGATGTTAAAGTTGTAATCTGACACT 741

QY 841 gtacttccacttgcagaagatttgaagatctgcaactgagagcagagctgactagt 900
|||||
Db 742 ATCACCTTTCATTTCCGTCTGACGACTATGATGACTTATGCCAGAGGACAGAGTCTCAGC 801

QY 901 ggaatcatgagatgaatgaatgaatgacacacataactgtgagctgtctgtgt 960
|||||
Db 802 GAATGGCTTTAAAGTTCATGAAAGCTTGGCCAAAAGTTCTCACTTGGCTGCTGTTACTA 861

QY 961 aaccaagctactctactaattatagagaggtctacttcaactgtctgtctagtgac 1020
|||||
Db 862 AACCGAGTGACCAACAAAGTTAGTGAAGGCTCTTCAACTAGCGTTGCTTTAGCGCAT 921

QY 1021 agctgtccactcactcagacagacagcggttgaatctgactggaatggaagacatgac 1080
|||||
Db 922 AGCTGTCTCATTCGTGACACCAACCGAGTCATCTGTAATGGAATGATGATGAGCTTAC 981

QY 1081 gacatctgtatagatctctcttacttcaactgagtgagcctgacagcagatgagtaggc 1140
|||||
Db 982 GCATATATCATTAAGTCCCTTCATCTTCACTTCACTTGGCTGCTTCACTACCTGATACCACT 1041

QY 1141 aaaggagtagaga 1154
|||||
Db 1042 AGAGGCTTAAGANA 1055

RESULT 5
AF029669
LOCUS Homo sapiens Rad51C (RAD51C) mRNA, complete cds.
DEFINITION AF029669
ACCESSION AF029669
VERSION AF029669.1 GI:2909800
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1295)
AUTHORS Dosanjh, M.K., Collins, D.W., Fan, W., Lennon, G.G., Albaladejo, J.S., Shen, Z., and Schild, D.
TITLE Isolation and characterization of RAD51C, a new human member of the RAD51 family of related genes
JOURNAL Nucleic Acids Res. 26 (5), 1179-1184 (1998)
MEDLINE 96136197
REFERENCE 2 (bases 1 to 1295)
AUTHORS Schild, D., Collins, D.W., and Dosanjh, M.K.
TITLE Direct Submission
JOURNAL Submitted (09-OCT-1997) Life Sciences, Lawrence Berkeley National Laboratory, Ms. 70A-1118, Berkeley, CA 94720, USA
LOCATION/Qualifiers

FEATURES
source 1..1295
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q:413.6cR from the top"
/cell_type="Leukocyte (mixed population)"

SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 122871)
AUTHORS Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Beilto, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnes, M.E., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., Vanden, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carreira, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Coppenhaver, G.P., Preuss, D., Nieman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 122871)
AUTHORS Lin, X.
TITLE Direct Submission
SUBMITTED (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
On Dec 17, 1999 this sequence version replaced gi:2583106.
The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tdb/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://arthur.epm.ornl.gov/pub/grail/>), GeneFinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/GENSCAN.html>), and NecPlantGene (<http://www.cds.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the GSH/Maxhu/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13J16, the ESSA group for sequencing clone F13B4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalek, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: atetigr.org.
Location/Qualifiers
1. 122871
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
misc_feature complement(<1..2952)
/note="Sequence from clone T14P1"
repeat_region complement(519..627)

mrna /rpt_family="(TAAA)n"
join(2227..2322,2480..2527,2650..2705,3098..3150,3250..3641)
/gene="At2g45170"
2227..3641
/gene="At2g45170"
/note="T14P1.2"
join(2488..2527,2650..2705,3098..3150,3250..3469)
/gene="At2g45170"
/codon_start=1
/product="putative microtubule-associated protein"
/protein_id="AA18602.1"
/db_xref="GI:6598366"
/translation="MNKSGIFKMDDEPKRKAERIRKRPDRIPVYKAEKSEVPNIDKKYIVPSDLTGQFVYVIRKRIKLSAEKAIFFVNDVLPPTGELMSVYEDKKD EDGELITYSGENTFGASST"
complement(2953..122871)
/note="Sequence from clone F4L23"
4282..4870
/gene="At2g45180"
4282..4870
/gene="At2g45180"
/note="F4L23.31"
4320..4724
/gene="At2g45180"
/codon_start=1
/product="putative proline-rich protein"
/protein_id="AA82643.1"
/db_xref="GI:2583134"
/translation="MSKALAVYALLITLNLFFFTVSTKCPPTTPPKPKPKSEKKAPAVKPCPTDITLKGADLGLVNVVSGSPKPCCTLLGLANLEAVGLCTALKAVNLGINLVNIDITLLNLVCGKRPVHGFCSS"
5379..5403
/rpt_family="POLY A"
complement(join(<6393..6470,6586..6663,7082..7157,7248..7296,7872..8004,8099..8260,8382..8495))
/gene="At2g45190"
complement(<6393..8495)
/gene="At2g45190"
/note="F4L23.30"
complement(join(6393..6470,6586..6663,7082..7157,7248..7296,7872..8004,8099..8260,8382..8495))
/gene="At2g45190"
/note="unknown protein"
/codon_start=1
/protein_id="AA82644.1"
/db_xref="GI:2583135"
/translation="MSKSSMSSPSSAVCSPPDHPSPSDHLCYVOCNFCOTILAVNPEYTSKRTVYRCCCTNLTLSVNRBSYVPSMLOLOLGRHSYFPDILIELDAPSNMNMAMNQHPTMNDIPSPMDLHOHEIPKAPVVRPEPKQRPSPATNRTKEIQRIKAGNDPISHREAFSAAKNMAHPHIHGLVPDNPVKTNMPQEGEDMMVMEGFA PAAANVGVPY"
10147..10201
/rpt_family="(GA)n"
12613..12646
/rpt_family="(CAT)n"
complement(join(15613..15729,15821..15950,16066..16137,16252..16341,16810..17006,17451..17647))
/gene="At2g45200"
complement(15613..17647)
/gene="At2g45200"
/note="F4L23.29"
complement(join(15632..15729,15821..15950,16066..16137,16252..16341,16810..17006,17451..17583))
/gene="At2g45200"
/codon_start=1
/product="putative cis-Golgi SNARE protein"
/protein_id="AA82642.1"
/db_xref="GI:2583133"
/translation="MTSSSLDDQESGMEELRRERARKIEGDLQVYKSSVYKLGARFTQG GYVDGSPVGRSGRWSKMEETOSLEKLDINDSMKSAASAPFTSVYQKLAKRH DILHEVYQEFRRKIGNINSLEHAEILSVRDIISSYKASGSMSPGVYLRBRASIHG SISHIDVIOQAQATRAVLGSGRSLFSDVGKVKNGDKFVPVIRGLGSIKKRRSDT

```

misc_feature
    complement(17986..18862)
    /note="molecular marker ML"
repeat_region
    /rpt_family="(TAAA)n"
    19741..>20315
    /gene="At2g45210"
    /note="At2g45210"
    /note="F4L23.28; member of auxin-induced protein multigene family"
    19827..20315
    /gene="At2g45210"
    /codon_start=1
    /product="putative auxin-induced protein"
    /protein_id="AAB82641.1"
    /db_xref="GI:2583132"
    /translation="MRKIGFRIGRRYRWIFKRTIORSYWRHISTQACMLMPLAKLSWGQRLKQSFRRSRRTSRSAIYPVDHKADPPRGHAIAYVGGKDGCHVLPDIVFNHPLFGLLRREAREKEKGFCHEGGITIPCLYSDEFRVKTRIASGSSSRVFPWGRHC RN"
mRNA
    complement(join(<22224..22909,23488..>24334))
    /gene="At2g45220"
    /complement(<22224..>24334)
    /note="F4L23.27"
    /complement(join(22224..22909,23488..24334))
    /gene="At2g45220"
    /codon_start=1
    /product="putative pectinesterase"
    /protein_id="AAB82640.1"
    /db_xref="GI:2583131"
    /translation="MAFRAYINFEVLICILVASTVSGYNOKDYKAWSCOTPNPCEYFLPNSNNEPKSESEFLKISMKLVDRALTAHTLPGKCRDTRREKAWMDCKLYDLTVSKLINFETMDPNVCSKLDQOTWLSLTALDLCRAGLELELQTDVILPMSNNVSNLICNTLAINKPENPTPEKDPSPSWKPGCRKLQOSTPDKNVAADSGNKTKKEAIDNAGSGSRVIVYKQGVYSENLEIRKKVNMGRDGIKTIITGSKSVGGTTFNSATVAAGDVGFARCIIFERNAGASNEQAVLARGSDLSYQCSFEAYODTLVHSNRFYRDQVYGVDFEFGNAAVLQNCNIFARRPSKNTITTAOGRSPDNQNTIILHNSRVTAAASDLRVLGSKTYLGRFWRVSRVFPKTSIDSLIDPRGMEFMGNALKTLEFYAFQNTGPGASISGRVTFGRFVLSASABAKFTYGTFLAGSNIPSSVPTTS GL"
CDS
    complement(join(<22224..22909,23488..>24334))
    /gene="At2g45220"
    /complement(<22224..>24334)
    /note="F4L23.27"
    /complement(join(22224..22909,23488..24334))
    /gene="At2g45220"
    /codon_start=1
    /product="putative pectinesterase"
    /protein_id="AAB82640.1"
    /db_xref="GI:2583131"
    /translation="MAFRAYINFEVLICILVASTVSGYNOKDYKAWSCOTPNPCEYFLPNSNNEPKSESEFLKISMKLVDRALTAHTLPGKCRDTRREKAWMDCKLYDLTVSKLINFETMDPNVCSKLDQOTWLSLTALDLCRAGLELELQTDVILPMSNNVSNLICNTLAINKPENPTPEKDPSPSWKPGCRKLQOSTPDKNVAADSGNKTKKEAIDNAGSGSRVIVYKQGVYSENLEIRKKVNMGRDGIKTIITGSKSVGGTTFNSATVAAGDVGFARCIIFERNAGASNEQAVLARGSDLSYQCSFEAYODTLVHSNRFYRDQVYGVDFEFGNAAVLQNCNIFARRPSKNTITTAOGRSPDNQNTIILHNSRVTAAASDLRVLGSKTYLGRFWRVSRVFPKTSIDSLIDPRGMEFMGNALKTLEFYAFQNTGPGASISGRVTFGRFVLSASABAKFTYGTFLAGSNIPSSVPTTS GL"
repeat_region
    23009..23038
    /rpt_family="(TA)n"
repeat_region
    28701..28741
    /rpt_family="POLY_A"
repeat_region
    complement(29132..29228)
    /rpt_family="POLY_A"

```

Query Match 6.9%; Score 101.4; DB 12; Length 122871;
 Best Local Similarity 59.0%; Pred. No. 7.5e-14;
 Matches 174; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```

585 tatagatacagaagcattcatggtgaagcgtgtaccagatgctgaagcgtgat 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49087 TACAGATACAGAAAGTACCTCATGAGGACGCTTTACAGATACAGAAAGTGTGT 49146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
645 tagggacatactggagcacttcgcacagcattgagaagccttccttcgtccaa 704
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49147 AGAGGACATCGAAGATACACAGATACATGATTAACATTTTAAAGCAAAATCAAGTACA 49206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 attacagccttgagcgttccctcgcgatcattacttcgcgatcagattacacga 764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49207 AATGAACACAGAAATATCTTAGAGAACATATCTACTTCGGTCTCGCAGTTACACCGA 49266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 acaaatgcagtcataactacatgagaaattccccaagagacataaagatgtgc 824
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49267 GCAATGCGCATTTGCTCAATCTTGAAGAAATCATCTGAAAGCAAGATGAGGTAT 49326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
825 agttatattgatagtgattacttcacacttcgcacagatttgaagatctgca 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49327 AATATTTTGAAGTTCCATCCGTTGATTGCTGCTACAAAGTGTATATGTGTGACA 49381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 11
 LMFICHR32_17
 WPCOMMENT

Sequence split into 28 fragments LOCUS LMFICHR32 Accession AL499622

Fragment Name	Begin	End	LOCUS	Accession
LMFICHR32_00	1	110000		
LMFICHR32_01	100001	210000		
LMFICHR32_02	200001	310000		
LMFICHR32_03	300001	410000		
LMFICHR32_04	400001	510000		
LMFICHR32_05	500001	610000		
LMFICHR32_06	600001	710000		
LMFICHR32_07	700001	810000		
LMFICHR32_08	800001	910000		
LMFICHR32_09	900001	1010000		
LMFICHR32_10	1000001	1110000		
LMFICHR32_11	1100001	1210000		
LMFICHR32_12	1200001	1310000		
LMFICHR32_13	1300001	1410000		
LMFICHR32_14	1400001	1510000		
LMFICHR32_15	1500001	1610000		
LMFICHR32_16	1600001	1710000		
LMFICHR32_17	1700001	1810000		
LMFICHR32_18	1800001	1910000		
LMFICHR32_19	1900001	2010000		
LMFICHR32_20	2000001	2110000		
LMFICHR32_21	2100001	2210000		
LMFICHR32_22	2200001	2310000		
LMFICHR32_23	2300001	2410000		
LMFICHR32_24	2400001	2510000		
LMFICHR32_25	2500001	2610000		
LMFICHR32_26	2600001	2710000		
LMFICHR32_27	2700001	2727709		

Continuation (18 of 28) of LMFICHR32 from base 1700001 (AL499622 Leishmania major chr

Query Match 5.9%; Score 87.2; DB 84; Length 110000;
 Best Local Similarity 61.4%; Pred. No. 2e-10;
 Matches 140; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

405 gaacacatactactgctgtgtgtgacccaatgacatactgtgtgaggaattactg 464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32936 GACCACAGTACCAACTCTCTGCGGAGCTCGACGCGTACTGTGAGTGCGGCTTCCGT 32995
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
465 caaagaagtactgagatggtggtgtcccaagggtgtgttaaaccaactcgtgattca 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32996 TGCGGCGGTACCGAGATCAGTGGGCCCTCGCGCTTGGCAAGACACACGCTTGATGCA 33055
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 actagcaatcaatgatacaatcccaatggaatggtgtgctgtgtggaagcaattta 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
33056 GCTAGCAGTTAGCTGTGCGCATGCGCACTAGAGTTGCGAGGAGTGGTGGCGCTGCTTT 33115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
33116 CGTTGATACGAGGAGGACGCTTGTGCTGAGCGCTTAGAGCAAAATGCG 33163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 12
 AE001035/C
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Archaeoglobus fulgidus.
 Archaeoglobus fulgidus
 Archaea: Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 Archaeoglobus.

REFERENCE
 1 (bases 1 to 10980)
 Klenk,H.P., Clayton,R.A., Tomb,J., White,O., Nelson,K.E.,
 Ketchum,K.A., Dodson,R.J., Gwinn,M., Hickey,E.K., Peterson,J.D.,
 Richardson,D.L., Kervavage,A.R., Graham,D.E., Kyrpides,N.C.,

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirtress, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Uitterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Moese, C.R. and Venter, J.C.

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*
Nature 390 (6658), 364-370 (1997)
98049343
2 (bases 1 to 10980)

Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E., Richardson, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Fleischmann, R.D., Kierlavage, A.R., Graham, D.E., Kyriakides, N.C., Kirtress, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Uitterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Moese, C.R. and Venter, J.C.

Direct Submission
Submitted (15-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA
In order to show the genes in ascending order on the genome, the origin of this version has been moved by TIGR to position 2093570 of the original version and the opposite strand is shown from the original version.
On Dec 16, 1997 this sequence version replaced gi:2649594.

COMMENT
FEATURES
source

1. 10980
Location/Qualifiers
/organism="Archaeoglobus fulgidus"
/db_xref="taxon:2234"
159..362
/gene="Af0992"
159..362
/gene="Af0992"
/note="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region Af0992"
/protein_id="AAB90251.1"
/db_xref="GI:2649605"
/translation="MLGRFRLGVLALVAVPAMANGEDI.RNTKYKPEKVRDQ
RGTIVNDITLDEARNLVAVWDVQVSLSEKRLKRLSNTSMKDEITAEINHTDL
INAKELIAETVEELKEMKARVWDVSLQKSLITIGVLDLLETVEKGELEN
FVEGRVIAEEEGDPTLLENNLDSYRHRMALEKIGEAKEVLEITPTQGEAMK
EYREAVKTAVOITKECVKDLREIILQINQYGADESELMQVEEVVEE"
complement(971..1984)
/gene="Af0993"
complement(971..1984)
/gene="Af0993"
/note="similar to GB:L77117 SP:Q49593 PID:1378034
PID:1591553 percent identity: 59.29; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA repair protein RAD51 (rada)"
/protein_id="AAB90248.1"
/db_xref="GI:2649602"
/translation="MSESNSEETKLTLEEDIPGVETARKLRLEAGSTIAVAVASP
SFLAVGITEENAVKTIQAAKLANIGFSGDKVLRRSYKRTITSGSDRLDELG
CGVETQATTEPFGEGSGKTQICHOANVQIPREGEGLSEYIITIDENPFRPRTI
QMAEAGLDGNEVLKNITVADATNSHMLDYNKELAEKIKKGRVRLITVDSLM
SHREAYVGRGLADROOKLNHLDLKKFGLNAAIVTNQVAVADVLEGGPTKP
VGGHVAHTATFRILYKKGKDLRIARLIDSPHLPEGEAIFRVERGIEDAEKDKK
RKK"
2040..2597
/gene="Af0994"

CDS

2040..2597
/gene="Af0994"
/note="similar to SP:P54168 PID:1256633 GB:AL009126
percent identity: 43.75; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB90247.1"
/db_xref="GI:2649601"
/translation="MSSHDEGHEVRIKRLARIAREGARLDVYKAAELHDIAGCP
NARKSAELARELKDHDAEVAVCHALSAHSQKPKTLEKTLISDAKDLAIG
AIGVARAFYISGEGRSIEETLKHFEKLLKLDLLTETARKIAEERRRLEIYRQ
IKKELDEFEEDQDAEDYVNGQHYK"
complement(2532..4487)
/gene="Af0995"
complement(2532..4487)
/gene="Af0995"
/note="similar to GB:L77117 PID:1591923 percent identity:
26.90; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB90246.1"
/db_xref="GI:2649600"
/translation="MKEANRLTYIGYRLFAEDIMKSKYFELERSLKAMISNPE
MYIATARMESLEIGIIGAVILIAFIVKVGVPFFENIRLPEPATMYLPRFVY
AGLAIGITLALYGLNLIFAIPTSVISDRKSKIRLIPALIMYSLSKGMISVQ
ITKALADNHEVGEVGEASRLIMEVEGKRLRLAVALVDPDNDKRELGLIT
IIDSQDITRFEERAEFPERARQOKSFLELGMATFYIAPAGPLITITOTV
MSVGOETNEVAMSLITYVPISEFMALIKLTPGEGSAPRLERITTYRRSEL
DSEKRLKERVYKRLMKLPRLIVENLYTFAFIPALVILIGLVNPEIP
GMDLKNWFEYDYDIFISLWVYGPALVEGLRGVAKYMRITPILSKLSANESG
MPTIRAITMAKDTSPILREIKAKADWDGSLSDALVRANRLKIEISRTWML
NEALSGVTEVLTAMISAKDNTASALLREBRNMEMYIYIAFVYIGIYIIS
TELSTLAESAQAQATCONOMOSVDYTLKRVNFMHAAVFGIFAGWAGVMSGSL
SAGAKSLIMLTATVYFVSRFKLI"
complement(4489..6198)
/gene="Af0996"
complement(4489..6198)
/gene="Af0996"
/note="similar to GB:L77117 SP:Q58310 PID:1591575 percent
identity: 41.71; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="type II secretion system protein (gspe-3)"
/protein_id="AAB90245.1"
/db_xref="GI:2649599"
/translation="MPKINRQAKFRGKQDESESEARQADYEKEALEAIEVYOK
EELIKVAREKKIIGLPKREKISVEEHGELVERISPPVGSIVBYELPEPYCKAFIT
NEEQRKRYGVNPRIDPYETEYVYGLSVLKRLDENFEDGKVAEVLKETPD
ISEVGELEDERSYKIMYIMRDFLYDKTTPMLDKMLEDISCNVRKYMVFHNRY
TNLRTNVIIPDDPELDSFVYVNLAKGKHISLEPVAIDMPGSRIOUOTLREVYDH
GSTFTIKRFEPEVPTDILIAKQTFSSBOAYVLIENKKSIFGAGTAAGTSMN
AISLFTPRRSKITYITEDTRIMLPHENKIPAVYRGLRHEGKAVAYDILRLARLRD
EYITIVGEVRREALTLQAMATGHTTSTLHADISGALHRLDENPVGIPRRMLRQ
IISQATYVGDRRVRNMEIATVIGDATTMLKRLSTVFOVOWSVADENAMGTSLAL
EIRIRGMSVREINELERRRRLVLEFMLEHNVRKRVSNIIHTYOTKPRDKIMEAIS
KEG"
complement(6259..7509)
/gene="Af0997"
complement(6259..7509)
/gene="Af0997"
/note="similar to GB:L77117 PID:1592173 percent identity:
44.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="N-ethylamine cholorohydrolase (tiza-2)"
/protein_id="AAB90244.1"
/db_xref="GI:2649598"
/translation="MHDLVIRNGLCFNCFEVCVGVGNRTTHVAKVEERGEIETD
AAGLVVPGCFNAHTAAMTLRLGVAEGPLRLRMLKVEVEARLDEDAVYVGTMLAC

[illegible]

VERSION	KEYWORDS
AEO00901.1	GI:2622486
SOURCE	Methanothermobacter thermautotrophicus.
ORGANISM	Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
REFERENCE	Methanothermobacter.
AUTHORS	1 (bases 1 to 13968) Smith,D.R., Doucellet-Stamm,L.A., Delonghery,C., Lee,H., Dubois,J., Alredge,T., Bashirzadeh,R., Blakely,D., Cook,R., Gilbert,R., Harrison,D., Hoang,L., Keagle,P., Lumm,W., Pochier,B., Qiu,D., Spadatoro,R., Vicaire,R., Wang,Y., Mierzbowski,J., Gibson,R., Jlani,N., Caruso,A., Bush,D., Reeve,J.N. et.al. Complete genome sequence of Methanobacterium thermoautotrophicum delta: functional analysis and comparative genomics J. Bacteriol. 179 (22), 7135-7155 (1997)
TITLE	98037514
JOURNAL	2 (bases 1 to 13968)
MEDLINE	Smith,D.R.
REFERENCE	Direct Submission
AUTHORS	Submitted (10-AUG-1997) Genomics and Technology Development, Genome
TITLE	Therapeutics Corporation, 100 Beaver Street, Waltham, MA
JOURNAL	02154-8448, USA
FEATURES	Location/Qualifiers
SOURCE	1..13968 /organism="Methanothermobacter thermautotrophicus" /strain="delta H" /db_xref="taxon:145262" /clone="MTH" 63..434 /gene="MTH1377" 63..434 /note="Function Code:14..01 - Unknown, Conserved protein; similar to, pir:LN:G64486 AC:G64486, p(1)=7E-10, pid=29%" /codon_start=1 /transl_table=1 /product="conserved protein" /protein_id="AAB85854.1" /db_xref="GI:2622487" /translation="MMKREYLNKTRGSLGAVRPARTFMSRFGMLRQVETGLVLE IPVGRGSGSIGHMFMLVPLDVFVVDGDMRWVSVLRPMQVNPARKPARVYIELRE GKIRDSGRVAGDTIEFRDLOS" complement(502..900) /gene="MTH1378" complement(502..900) /gene="MTH1379" complement(502..900) /gene="MTH1378" /note="Function Code:9..10 - Metabolism of Cofactors and Vitamins, Porphyrin and chlorophyll metabolism ; similar to, sp:LN:CPCE_FREDI AC:P07126:P05520, p(1)=1.7E-09, pid=55%" /codon_start=1 /transl_table=1 /product="Phycocyanin alpha phycocyanobilin lyase Cpce" /protein_id="AAB85855.1" /db_xref="GI:2622488" /translation="MFEWEGKRIDFLLEELRPMPVYREDAVELLAEPDPAVGPILIE EALDDEYHVEAAALALAFRDDMAVEPEHEHSDKPGYRACALALGILGDEDSI PDLEILLDESPMWRYAAEVAISEIRKRAA" complement(1027..11737) /gene="MTH1379" complement(1027..11737) /gene="MTH1379" complement(1027..11737) /note="Function Code:14..01 - Unknown, Conserved protein; similar to, sp:LN:Y208_METUA AC:Q57661, p(1)=1.2E-51, pid=48%" /codon_start=1 /transl_table=1 /product="conserved protein (contains ferredoxin domain)" /protein_id="AAB85856.1" /db_xref="GI:2622489" /translation="MRIAMFTGAGHLLSEVALMEAVEGCHEVTLLSGAAEEVLR MYGLIEPKRSIGCYRELIDDEQGSFPTGTLGMSGRDILVSPVTSNTAKVYH GIADLTITNNAQAQKGGVPYCYCPVLDDESDVETVLPKSLIEELCKGCEPCALAAAC

```

gene
CDS
/feature="PEDIAVPGVEIRLLSCNGCAGCTACPHGAVSGRIITIHREVDIRNTARLAAMDGI
RVFAKPGILEMMAHSI"
/feature="complement(1738..3024)"
/feature="gene="MTH1380"
/feature="complement(1738..3024)"
/feature="gene="MTH1380"
/feature="note="Function Code:5.02 - L-Amino Acid Metabolism,
glycine--serine and threonine metabolism ; similar to,
sp:LN:GLYA.METH AC:P50436, p()=5..9E-203, pid=85%"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="AAB85857.1"
/db_xref="GI:2622490"
/translation="MEMIFVSNODYTEKIRQLMKDHSNMISSINLIASENITSRY
KEALSDSHRYAEGLPGERLYEGCRYIDEIEELTELKRIEFAEHANVQPTSGVA
NACPFATAEVGDPTMAEMVEPYGGHISHARYSAGVGRFOYTHPPDENNDIMADP
KKITIEVKPRITILEGSLFLFPHRYEALAEAEYCARINTDGAHVIGLITAGTFOOP
LREGADMLVSTHTFPGPGGILLCRELAADIDEAVFGLVSNHHLHVAIGLIAT
AEMLEFGAEVAAOTINNARKLAENLHELGFVNLCEHDFESHQVNDVSDIGRAEI
SKRLANNIINKNLPMVDVNRSDPDSGIRIGIQETIRGMKESESEVAYEYIKRVY
MDGKVRDEVAEFMSSTYRVHAFEDSEAYKYMEO"
/feature="complement(3192..5201)"
/feature="gene="MTH1381"
/feature="complement(3192..5201)"
/feature="note="Function Code:10.08 - Metabolism of Macromolecules,
protein translation and modification ; similar to,
pir:LN:S48720 AC:S48720, p()=0, pid=88%"
/codon_start=1
/transl_table=11
/product="heterodisulfide reductase, subunit A"
/protein_id="AAB85858.1"
/db_xref="GI:2622491"
/translation="MLDTPMOGDLMAEKKEETMEPRIGVYVCHGNGIVDIEA
VRDYAKLPNVVSKDYKYCSDPGOLEIKDKIEGINRVYVACSPRIHEPFRFC
VEEGLNGLFEPANLPRDSDWVHMDDEGATEKAKLYRMAVAKALBLESKYS
VDKRLVIGGVAGIOTRLDLADMGFTTYVEKRPISGRGQIDKTFPIIDCSMCLL
APKRVYDVKHNDIELITYAEVKEVDYIGNFKVRIEKKPRYIDEDLCTGCGSCVECP
IEMPVDEGIGKAVYIPPOAVPLCATIDKDCIECMCDEICERGAHVHQBEP
EIELEVGIIIVATGADAVDPTKELEYGRHTNYTGIELERMINASGPTGKVIKS
DGEKRVAFIHGCSRDEOIGKPYSCVCCMIKNAOLIKDKMPTDEVLITMDIR
AFGKEFEERYKSOEKYGIKIRGRPAELIENPDILTJVRSEDFLLKRYEYDQDMY
IGCVLVPRESSEKRLQRTIGLSKADGFLMEHPRKLPDITLTDGVYLAGVQKQDIP
DAVQAQSGAARAAIIPVYKGEVEIEPIVAVTSDVCGCEVCIELCPFGAISTEEGA
NNVALKGCCTCAACPGSGAMDQOHEFTBOIMAOIRALMEQAK"
/feature="complement(5353..6261)"
/feature="gene="MTH1382"
/feature="complement(5353..6261)"
/feature="note="Function Code:14.01 - Unknown, Conserved protein;
similar to, pir:LN:F64428 AC:F64428, p()=3..7E-20, pid=24%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85859.1"
/db_xref="GI:2622492"
/translation="MIFMGSTEVLSVLVLLVGLVGLGEEDASTLNRYVINV
AIPSLIFSTRADLSGLDVLIPVICMTIGTIGTILAYIMARGLDSDGKMGVLV
AAAMNSGELGYPVTAIGFSGEGLVRAIFDGTGTLMTSGILLSHISGEGSRIMK
RAYFPPLMAFLVITRNKGLPTGINGITIGISGAAPLIMISLULITNFRRLRS
VADATVSGRLKILISPLMAAGISTVLVRLGNTFVLEASMPASMLAAVALAENDID
VDLVSSCIFNSTILSVSLPLMSVVL"
/feature="complement(6276..7211)"
/feature="gene="MTH1383"
/feature="complement(6276..7211)"
/feature="note="Function Code:10.09 - Metabolism of Macromolecules,
DNA replication--modification--repair--and recombination ;
similar to, sp:LN:RADA_SULSO AC:Q55075, p()=6..6E-85,
pid=51%"
/codon_start=1
/transl_table=11
/feature="product="DNA repair protein Rada"
/feature="protein_id="AAB85860.1"
/feature="db_xref="GI:2622493"
/feature="translation="MVEIDLPVYGAKTKTAOKLRDAGFGDMRLATATKELSVKAEIG
EGVAEYIEAKARAEEKIDFETAPDWERKKDVRGRTGSKALDELIGGITEFOATTEV
EGEGSGKSQLSHELAVTQLPFERGDLAEAVFIDETFTFERPERIEQIYANFELDELE
EVLNKTHIAFAENSSHQILMAEKVNELIOBKRIIRLIVDSLTAFHRAEYVORELAT
ROOKLNDHHTLTONIANTYNAAVFVNOVQARDAFSGSPTRAGIHVGHVGHAAVYRIW
LRKGLAGKRIARLVSDPHLPEGECVFKITTAGIVD"
/feature="complement(7264..7809)"
/feature="gene="MTH1384"
/feature="complement(7264..7809)"
/feature="note="Function Code:10.09 - Metabolism of Macromolecules,
DNA replication--modification--repair--and recombination ;
similar to, pir:LN:F64444 AC:F64444, p()=0.0000042,
pid=21%"
/codon_start=1
/transl_table=11
/product="replication factor A related protein"
/protein_id="AAB85861.1"
/db_xref="GI:2622494"
/translation="MKLSIGNSRIEIPASERDLGEPFDEQEMHYPRDADDED
SRNVLIEGLIELMSGRRLISIKPSCNRELSIDENICNFGELVDEPRYLLMTPGRI
MDDTGEVMITFEGERAESILEMTTDEVNINQSADESALAEERVDLNGVTVVIGNA
DMQVSEBELRFPKRVKKEL"
/feature="complement(7775..9643)"
/feature="gene="MTH1385"
/feature="complement(7775..9643)"
/feature="note="Function Code:10.09 - Metabolism of Macromolecules,
DNA replication--modification--repair--and recombination ;
similar to, pir:LN:F64444 AC:F64444, p()=3..7E-31, pid=45%"
/codon_start=1
/transl_table=11
/product="replication factor A related protein"
/protein_id="AAB85862.1"
/db_xref="GI:2622495"
/translation="MKELKREYERIKRDISPEFEELIEKKKEIGDIFMDLITIA
SYVVDILIEKNTMLSEKPEHNDITSLKEBAEPYVGRVAKISSPRTFTKRREG
KLANVLIDDTGDLRAVFTNENIKLKREDDVIRIKVNRIGSGFERGKEHLMPSG
TYVLDVPEPEPEYREITPTGIDVEDDEVNVAIRATGVSVEYFREDGREGFIS
LOIMDQSTPTTYTNNVNDLVBELGKSGDAVKILMAOPRRDRKVTILTHSLTRYV
PGRTDYPERFEELVKTGDLHEKNVTVGLVTVKNDPVFERNDCTGSVSIETADV
TGSARVTLWDEDMRIKINGDIIRISGANVEDDNOISRIINTNRTITLNPESDA
LKALVEYERQMPKISEILMEDEGEVDVAGIFSLSDREFREDGTGIVSME
LADETKIRISLWDEKAERPMNIGDAVRIENARIRGLVSYFSRGTRTIYVLPED
MEDPSEFEELKMTYOTKRIADLEEDRIIRIIRIIVYDLFEPRERODGCGIVRTA
EPADDTGSTRASIMDDAAEKPLSIDGPVYIENPRVYFRDDMGGRLETOHKKQFD"
/feature="9947..11206"
/feature="gene"
Query Match
Best Local Similarity 5.1%; Score 74.6; DB 1; Length 13968;
Matches 140; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
gene
Query Match
Best Local Similarity 5.1%; Score 74.6; DB 1; Length 13968;
Matches 140; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```


Db 89267 CAGGAACCTTGACGAGCTGCTGGCGCGGTATAGAGACTAAGACTATATAACGAGCTTT 89326
QY 486 tggcgtcccaagggttggtaaaactcaactgggattcaactagcaatcaatgtaaat 545
Db 89327 CGGTGAGTTGCGGAGCGGCAGACTCAGATATGCCATCAGCTATCTGTTAACGTCACGCT 89386
QY 546 cccagtggaatgtggtggccttggtggaagcagttatagatcacagggcagttt 605
Db 89387 CCCGGAAGACAGGAGGCTCGAGGGTAAAGCGGTATACGTAGATACCGAGGCACTT 89446
QY 606 catggttgaacgtgtctaccagaattgctgaag 638
Db 89447 CAGGTGGAGAGGATAGAGCAGATGGCTGCTGG 89479

Search completed: October 2, 2001, 05:54:16
Job time: 3732 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time: 11.19 Seconds
(without alignments)
900.010 Million cell updates/sec

Title: US-09-537-654-2

Perfect score: 1530
Sequence: 1 MGDQSGSRNGPQOKYVSGAQ.....VTGKGRDVAESNHRARAT 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	576.5	37.7	376	1	R51C_HUMAN
2	387.5	25.3	324	1	RADA_SUIISO
3	354.5	23.0	337	1	RADA_ARCFU
4	352.5	23.0	311	1	RADA_METH
5	337	22.0	334	1	DMC1_YEAST
6	323	21.1	350	1	R51B_HUMAN
7	314.5	20.6	340	1	DMC1_MOUSE
8	313	20.5	340	1	DMC1_HUMAN
9	313	20.5	349	1	DMC1_LILLO
10	313	20.5	350	1	R51B_MOUSE
11	312	20.4	345	1	DMC1_SOYBN
12	312	20.4	352	1	RADA_METJA
13	304	19.9	324	1	DLH1_CANAL
14	299.5	19.6	339	1	RAS1_USRMA
15	297.5	19.4	332	1	DMC1_SCHPO
16	290	19.0	342	1	DMC1_ARATH
17	287.5	18.8	339	1	RAS1_CHICK
18	287.5	18.8	333	1	RADA_HALVO
19	284.5	18.6	339	1	RAS1_CRGR
20	282.5	18.5	336	1	R511_XENLA
21	281.5	18.4	339	1	RAS1_HUMAN
22	280.5	18.3	336	1	R512_XENLA
23	280.5	18.3	339	1	RAS1_MOUSE
24	277.5	18.1	400	1	RAS1_YEAST
25	276.5	18.1	339	1	RAS1_RABIT
26	276	18.0	365	1	RAS1_SCHPO
27	275.5	18.0	336	1	RAS1_DROME
28	272.5	17.8	342	1	RAS1_LICES
29	241.5	15.8	460	1	RAS1_YEAST
30	217.5	14.2	328	1	R51D_HUMAN
31	209	13.7	329	1	R51D_MOUSE
32	206	13.5	221	1	R51L_ARCFU
33	199	13.0	212	1	R51L_METJA

34	199	13.0	228	1	R51L_PYRHO	057859	pyrococcus
35	189	12.4	234	1	R51L_METH	027728	methanobact
36	174.5	11.4	216	1	R51L_METMP	050248	methanococc
37	164.5	10.8	350	1	RH55_SCHPO	014129	schizosacch
38	134.5	8.8	452	1	RADA_LISMO	048761	listeria mo
39	133	8.7	458	1	RADA_BACSU	037572	bacillus su
40	127.5	8.3	406	1	RAS5_YEAST	038953	saccharomyc
41	125.5	8.2	458	1	RADA_HAEIN	045266	haemophilus
42	118	7.7	343	1	RECA_XANCP	060101	xanthomonas
43	117.5	7.7	448	1	RADA_RICPR	092604	ricketsia
44	117.5	7.7	453	1	RADA_PSEAE	096963	pseudomonas
45	117.5	7.7	480	1	RADA_MYCTU	053570	mycobacteri

ALIGNMENTS

RESULT	1	STANDARD	PRT	376 AA.
ID	R51C_HUMAN			
AC	043502			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	DNA REPAIR PROTEIN RAD51 HOMOLOG 3.			
GN	RAD51L2 OR RAD51C.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-98136197; PubMed-9469824;			
RA	Dosenjib M.K., Collins D.W., Fan W., Lennon G.G., Albalade J.S., Shen Z.,			
RA	Schild D.;			
RT	Isolation and characterization of RAD51C, a new human member of the			
RT	RAD51 family of related genes.*			
RL	Nucleic Acids Res. 26:1179-1184(1998).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES, WITH			
CC	HIGHEST EXPRESSION IN TESTIS, HEART MUSCLE, SPLEEN AND PROSTATE.			
CC	-1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO			
CC	PROKARYOTIC RECA PROTEIN.			
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: AF029669; AAC39604.1; -			
DR	MIM: 602774; -			
DR	PROSITE: PS50162; RECA_2; 1.			
DR	DNA-binding; ATP-binding; Nuclear protein.			
FT	NP_BIND 125 132 ATP (POTENTIAL).			
FT	SEQUENCE 376 AA; 42189 MW; 3AAAD3C1C0851E0 CRC64;			

Query Match 37.7%; Score 576.5; DB 1; Length 376;
Best local Similarity 43.9%; Pred. No. 2.2e-42;
Matches 125; Conservative 47; Mismatches 98; Indels 15; Gaps 4;

QY	22	AWMFEDELISOKITFGSDLDIIIGGHNCKEVEIGSPGVGTQGLGIQIAINQIPV	81
DB	87	ALLELDQEHQGGITIFGCALDDILGGVPLMKTTETICAPVGTQLCMQALVDDQD-	145
QY	82	EC-GGLGKAVYIDEGSPMERVVOIAEGCTRDI-LEHFPSSHKSSSVQKQLQPERFL	139
DB	146	ECGGVAGAEVFIIDEGSPMDVRVVDLATACIQHQLIAEKHKGSEHRALEDFLDNTL	205
QY	140	ADIIYRICSYTEQIAVINIMKEFLREHKDVRIIIVDSVTFHFRODFEDLALRTVLSGL	199

```

Db 206 SHIYRCRDYETLAAQVYLLPDLSEHSKVRVLYVDGIAFFERHDLDDLSTRTLNGL 265
Qy 200 SLKMTIATYINLAVVLLNOVTTKTEGFSQTLTALGDSMSHSCNRLILHNGMERARH 259
Db 266 AQQMISLANHNRHLAVLTITNQMTTKIDRNQALLVPALGSGWGHAAATRIILFHWDKORLAT 325
Qy 260 LDKSPSLPVASAPYAVTGGIRDAV-----SSNHKRRAR 292
Db 326 LKSPSQKECTYVLFQIKPGFRDYVTSACSLQTESLSTRKRSR 370

RESULT 2
RADA_SULSO STANDARD: PRT: 324 AA.
AC 055075;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA REPAIR PROTEIN RADA.
GN RADA.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96251095; PubMed=8668545;
RA Sandler S.J., Satlin L.H., Samra H.S., Clark A.J.;
RA "reca"-like genes from three archaean species with putative protein
RT products similar to Rad51 and Dmc1 proteins of the yeast
RL Saccharomyces cerevisiae."
CC Nucleic Acids Res. 24:2125-2132(1996).
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPAIR AND IN HOMOLOGOUS
RECOMBINATION AND REQUIRED FOR MEIOTIC RECOMBINATION.
CC -1- SIMILARITY: STRONG TO EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U45310; AAC4123.1; -
DR HSSP: P03017; 2REB.
DR PROSITE: PS50162; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
KW DNA damage; DNA recombination; ATP-binding; DNA-binding.
FT NP_BIND 114 121 ATP (POTENTIAL).
SQ SEQUENCE 324 AA; 35866 MW; B08CA699AA40A88 CRC64;

Query Match 25.3%; Score 387.5; DB 1; Length 324;
Best Local Similarity 34.2%; Pred. No. 3.5e-26;
Matches 97; Conservative 61; Mismatches 93; Indels 33; Gaps 5;

```

```

Qy 241 HSCITNRLILHWN-GENERYAHLDKSPSLPVASAPYAVTGGIRDA 293
Db 279 HVPGRITQLKSRNRRILARVADPHLEGEVYFALTEEGIRDA 322

RESULT 3
RADA_ARCFU STANDARD: PRT: 337 AA.
ID RADA_ARCFU
AC 029269;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA REPAIR PROTEIN RADA.
GN RADA OR AF0993.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=96049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Kelchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervatage A.R., Graham D.E., Kyriades N.C.,
RA Flischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Goodyear J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Styles S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPAIR AND IN HOMOLOGOUS
RECOMBINATION AND REQUIRED FOR MEIOTIC RECOMBINATION.
CC -1- SIMILARITY: STRONG TO EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001035; AAB90248.1; -
DR TIGR: AF0993; -
DR PROSITE: PS50162; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
KW DNA damage; DNA recombination; ATP-binding; DNA-binding.
FT NP_BIND 115 122 ATP (POTENTIAL).
SQ SEQUENCE 337 AA; 37162 MW; 7888592BAA3BD43 CRC64;

Query Match 23.2%; Score 354.5; DB 1; Length 337;
Best Local Similarity 32.4%; Pred. No. 2.6e-23;
Matches 95; Conservative 50; Mismatches 105; Indels 43; Gaps 6;

```


Db 171 NEVLAKNIYAQAAYNSNHQMLVLDNAKELAKBGRPVRLIIVDSLSHFRFAEYVGRGT 230
 QY 189 LALFRFVLSGLSLKMLKRIKATYNLAIVLLNOYTKFTGEGSFQTLALGDSW-----S 240
 Db 231 LADROCKLNRLHDLMLKRESELYNAAIVTNOVMAR-----PDVLFSDPTKPVGCHIVA 283
 QY 241 HSCNTNRLILHMGNE-RVAHLDKSPSLPVASAPVAATGKIGINDAVSNHKKR 292
 Db 284 HTATFRYILKKGKDLRIARLIDSPHLEGEAIFVTEGIDAEKDKKKRK 336

RESULT 4
 RADA_METTH STANDARD: PRT: 311 AA.

AC 027436;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RADA PROTEIN.
 GN RADA OR MTH1383.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea: Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 CC Methanobacter.
 ON NCBI_TaxID=145262;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RX MEDLINE=96037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
 RA Spadafora R., Viare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Nollig J., Reese J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -I- FUNCTION: PROBABLY INVOLVED IN DNA REPAIR AND IN HOMOLOGOUS
 RECOMBINATION AND REQUIRED FOR MEIOTIC RECOMBINATION.
 CC -I- SIMILARITY: STRONG TO EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
 PROKARYOTIC RECA PROTEIN.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 DR EMBL: AE000901; AA85860.1;
 DR PROSITE: PS50162; RECA_2; 1.
 DR PROSITE: PS50163; RECA_3; 1.
 KM DNA damage: DNA recombination: ATP-binding; DNA-binding.
 FT NP BIND 104 111 ATP (POTENTIAL).
 SQ SEQUENCE 311 AA; 34125 MM; 3A0DF609A8785DB CRC64;

Query Match 23.0%; Score 352.5; DB 1; Length 311;
 Best Local Similarity 33.6%; Pred. No. 3.5e-23;
 Matches 96; Conservative 46; Mismatches 111; Indels 33; Gaps 5;

QY 10 GPQKYYVSGA-----QNAWMFSDLSQKHITSGSDNDILGGSHCKEYELGV 61
 Db 46 GVAEYVIEAARAERKIDIEADVVERKKDVGRTTSGKALDELGGIETQATTEVGE 105
 QY 62 PGVKTQGLGDLAINVQJPEVGGGAGKAVYIDEGSFMEVRYQIAEGCTRIDLEHPPH 121
 Db 106 FGSGKSQLSHELAVTVQPEPGRGDAEAVPIDEFPRPRRIQAINAF----- 155
 QY 122 SHEKSSVOKOLOPRLADIIYFRICSTQIAIVINMEKFLREHKDVRIVITDSVTFH 181

Db 156 -----ELDLEEVNLKIHIAFNSSHQILMAEKNELIQEKNIRLIVDSLTAH 205
 QY 182 PROPE-----EDLALFRVLSGLSLKMLKRIKATYNLAIVLLNOYTKFTGEGSFQTLALGDS 238
 Db 206 FRABYVREALATFQOKLNHLHTLQNIANTYNAVAFTNOVAPDPAPFGSPKAIGH 265
 QY 239 -WSHSCNTNRLILHWN-GNERYAHLDKSPSLPVASAPVAATGKIGIRD 282
 Db 266 VLGHAATYRIWLKKGAGKRIARLIVDSPHLPGEQCVKITTAGIVD 311

RESULT 5
 DMCI_YEAST STANDARD: PRT: 334 AA.

AC P25453;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE MEIOTIC RECOMBINATION PROTEIN DMCI.
 GN DMCI OR ISC2 OR YER179W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SK1;
 RX MEDLINE=92257586; PubMed=1581960;
 RA Bishop D.K., Park D., Xu L., Kleckner N.;
 RT "DMCI: a meiosis-specific yeast homolog of E. coli recA required for
 RT recombination, synaptonemal complex formation, and cell cycle
 RT progression.";
 RL Cell 69:439-456(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SK1;
 RX MEDLINE=93204898; PubMed=8455558;
 RA Kobayashi T., Hotta Y., Tabata S.;
 RT "Isolation and characterization of a yeast gene that is homologous
 RT with a meiosis-specific cDNA from a plant.";
 RL Mol. Gen. Genet. 237:225-232(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
 RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.

CC
 CC -I- FUNCTION: REQUIRED FOR MEIOTIC RECOMBINATION, SYNAPTONEMAL COMPLEX
 FORMATION AND CELL CYCLE PROGRESSION.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.
 CC -I- DEVELOPMENTAL STAGE: EXPRESSED 2.5 HOURS AFTER INDUCTION OF
 MEIOSIS.
 CC -I- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
 PROKARYOTIC RECA PROTEIN.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).


```

RP SEQUENCE FROM N.A.
RC STRAIN-129/SV; TISSUE=Testis;
RX MEDLINE=96173646; PubMed=8602360;
RA Habu T., Taki T., West A., Nishimune Y., Morita T.;
RT "The mouse and human homologs of DMCL, the yeast meiosis-specific
RT homologous recombination gene, have a common unique form of exon-
RT skipped transcript in meiosis."
RL Nucleic Acids Res. 24:470-477(1996).
CC -1- FUNCTION: MAY PARTICIPATE IN MEIOTIC RECOMBINATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D58419; BAA09590.1; -
DR EMBL: D64107; BAA10969.1; -
DR HSSP: P03017; 2REB.
DR MIM: 602721; -
DR MGI: 105393; Dmclh.
DR PROSITE: PS50162; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
DR Meiosis; Cell cycle; ATP-binding; Nuclear protein.
KW NP_BIND 126 133 ATP (POTENTIAL).
FT SEQUENCE 340 AA; 37821 MW; 23EAB08D6E68637 CRC64;
SQ
Query Match 20.6%; Score 314.5; DB 1; Length 340;
Best Local Similarity 33.2%; Pred. No. 7.3e-20;
Matches 87; Conservative 43; Mismatches 97; Indels 35; Gaps 7;
QY 34 HITGSGDLNDILGGGIRCKEYETIGVGVGKTQIGIOLAINVOIPVECGLGKAVYI 93
DB 100 HITGSGQFEDKLLGGGIESMAITEAFGEFTGTQLSHTLCVTAQLPGTGTSGGKIIFI 159
QY 94 DTGSGFMERYVQIAEGCIRDLIEHPHSHKSSSVQKQLOPERFLADYIFRICSTEQ 153
DB 160 DTEMFERDR-----LRDIADRFNVDR-----EAVLDNVYARAVTSEHQ 199
QY 154 IAVIYVMEKFLREHKDVRIIVDSVTFHFRDPE--DLAFLRVVSGSLKIMTAKT 209
DB 200 MELLDYVAAKHEHAGIFKLLIDISIMALFRVDSGKGLAEKQKLAQMUSRLQKISEE 259
QY 210 YNLAVVLLNQVTTKFTGEGSFQTLAL-----GDSWSHSCNRLILH-WNGNERVAHLD 261
DB 260 YNVAVFVNTQMT--ADPGATMTFOADPKPKPIGHIILAHASTRTISLRKGGELRIAKIY 316
QY 262 KPSPLPVAASAPYAVTGKIRDA 283
DB 317 DSPMPENENATFAITAGIGDA 338
Db
RESULT 8
DMCL_HUMAN STANDARD; PRT; 340 AA.
AC 014565; 099498;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MEIOTIC RECOMBINATION PROTEIN DMCL/LIM15 HOMOLOG.
GN DMCL OR DMCLH OR LIM15.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Testis;
RX MEDLINE=96127532; PubMed=8590282;
RA Sato S., Seki N., Hotta Y., Tabeta S.;
RT "Expression profiles of a human gene identified as a structural
RT homologue of meiosis-specific reca-like genes."
RL DNA Res. 2:183-185(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=96173646; PubMed=8602360;
RA Habu T., Taki T., West A., Nishimune Y., Morita T.;
RT "The mouse and human homologs of DMCL, the yeast meiosis-specific
RT homologous recombination gene, have a common unique form of exon-
RT skipped transcript in meiosis."
RL Nucleic Acids Res. 24:470-477(1996).
CC -1- FUNCTION: MAY PARTICIPATE IN MEIOTIC RECOMBINATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D53882; BAA09932.1; -
DR EMBL: D64108; BAA10970.1; -
DR HSSP: P03017; 2REB.
DR MIM: 602721; -
DR PROSITE: PS50162; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
DR Meiosis; Cell cycle; ATP-binding; Nuclear protein.
KW NP_BIND 126 133 ATP (POTENTIAL).
FT CONFLICT 37 37 I -> N (IN REF. 2).
FT CONFLICT 183 183 P -> A (IN REF. 2).
FT SEQUENCE 340 AA; 37707 MW; EBBBDE57EB440402 CRC64;
SQ
Query Match 20.5%; Score 313; DB 1; Length 340;
Best Local Similarity 32.7%; Pred. No. 9.8e-20;
Matches 89; Conservative 45; Mismatches 98; Indels 40; Gaps 8;
QY 29 ELSQK-----HITGSGDLNDILGGGIRCKEYETIGVGVGKTQIGIOLAINVOIPVE 83
DB 90 EYSEKRRWFHITGSGQFEDKLLGGGIESMAITEAFGEFTGTQLSHTLCVTAQLPGAG 149
QY 84 GGLGKAVYIDTGESFMERYVQIAGCIRDLIEHPHSHKSSSVQKQLOPERFLADYI 143
DB 150 GYGGGKIIFIDIENTRPRDR-----LRDIADRFNVDR-----VLDNVL 189
QY 144 YFRICSTEQIAVINYM-EKFLREHKDVRIIVDSVTFHFRDPE--DLAFLRVVSGSL 199
DB 190 YARAVTSEHQMELLDYVAAKHEHAGIFKLLIDISIMALFRVDSGKGLAEKQKLAQM 249
QY 200 SLKIMIAKTYNNAVYLLNQVTTKFTGEGSFQTLAL-----GDSWSHSCNRLILH-W 251
DB 250 LSRLOKISEYVNAVFTQMT--ADPGATMTFOADPKPKPIGHIILAHASTRTISLRKG 306
QY 252 NGNERVAHLDKPSPLPVAASAPYAVTGKIRDA 283
DB 307 RGEIRIAKIYDSPMPENENATFAITAGIGDA 338
Db
RESULT 9
DMCL_LILLO STANDARD; PRT; 349 AA.
AC P37384;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)

```

```

DE MEIOTIC RECOMBINATION PROTEIN DMC1 HOMOLOG.
GN L1M15.
OS Liliun longiflorum (Trumpet lily).
OC Eukaryota, Viridiplantae, Embryophyta, Tracheophyta, Spermatophyta,
OC Magnoliophyta, Liliopsida, Liliales, Liliaceae, Liliun.
OX NCBI_TaxID=4690;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=flower buds;
RX MEDLINE=96051386; PubMed=7584025;
RA Kobayashi T., Kobayashi E., Sato S., Hotta Y., Miyajima N., Tanaka A.,
RT Tabata S.;
RT "Characterization of cDNAs induced in meiotic prophase in lily
microspocytes."
RL DNA Res. 1:15-26(1994).
CC - FUNCTION: MAY PARTICIPATE IN MEIOTIC RECOMBINATION.
CC - SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC - SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC DR EMBL: D21821; BAA04845.1; -.
DR PIR: JC2214; JC2214.
DR PROSITE: PS50163; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
KW Meiosis; Cell cycle; ATP-binding; Nuclear protein.
KW NP_BIND 138 145 ATP (POTENTIAL).
SQ SEQUENCE 349 AA; 38272 MW; CDA444A116DBEC6 CRC64;

Query Match 20.5%; Score 313; DB 1; Length 349;
Best Local Similarity 33.3%; Pred. No. 1e-19;
Matches 92; Conservative 48; Mismatches 104; Indels 32; Gaps 7;

QY 15 YVSGAONAMDMFSDLSOKHITPGSGDLNDLIGGICKKEVTEIGGVPGVGTQIGIOLA 74
   1::: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 97 YITGS---DVLLKRSYVIRITTSQADDELGGIETLQITEARGEFRSGKTQIAHTE 152
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

QY 75 INVQIPVECGGJGKRAVYIDTEGSGFENVEVYQIAECIFDLEHFRPHSKSSVYOKQLO 134
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 153 VSTQLPVSMHGNGKAYVYDTEGTFRPPRIYPIAEFGMD-----ASAV----- 196
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

QY 135 PERFLADIYFFICSTGEQIAVINMEKRLBHKDRIYIDSVTFHFRODPE--DIAL 191
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 197 ----LDNIITAYATYEHQYINLLALAAKMS-EPFRLLIVDSVIALFRVDESGRELAE 251
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

QY 192 KTRVYLSGLSLKMTAKTAYINLAVLINQVTTYFTBSF--QTLALGDSWHSCHNRLI 248
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 252 RQDKLAQMSRLTKIAEEFNVAAYVMNOVIADPGGGMFTSDPKKPKAGHYLAHAATVRLM 311
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

QY 249 LHWNGENERYAHLDKSPSLPVASAPAYVNGKGRDA 283
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 312 LKRGKGGRVCKIFDAPNLPESAEVAFQITPGGVADA 347
   1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

RESULT 10
R51L MOUSE STANDARD; PRT; 350 AA.
AC 035719;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA REPAIR PROTEIN RAD51 HOMOLOG 2.
GN RAD51L1 OR RAD51B OR REC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97352813; PubMed=9207106;
RA Rice M.C., Smith S.T., Bullrich F., Hayre P., Kmiec E.B.;
RT "Isolation of human and mouse genes based on homology to REC2, a
RT recombinational repair gene from the fungus *Ustilago maydis*.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:7417-7422(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE RANGE OF TISSUES.
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/announce/
CC or_send_an_email_to_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch)).
CC -----
CC EMBL; U92068; AAB63359.1; -.
CC MGD; MGI:1099436; Rad51l.
CC PROSITE; PS50162; RECA_2: 1.
CC DNA-binding; ATP-binding; Nuclear protein.
CC NP_BIND 108 115 ATP (POTENTIAL).
CC SEQUENCE 350 AA; 38137 MW; A57F347CA9CE6D65 CMC64;
SQ
FT
SQ
Query Match 20.5%; Score 313; DB 1; Length 350;
Best Local Similarity 32.1%; Pred. No. 1e-19;
Matches 100; Conservative 46; Mismatches 98; Indels 68; Gaps 13;
QY 20 GNAWDM---FSDLSOKHTTSGDPLNDLIGCGHCKEYTEIGVPGYGTQDGLGQLAIN 76
DB 65 QYAVELKTRSRASHLESPAFSLTLCALDEAHGVCPCGSLTELTGPGCGCKQFCIMMSVL 124
QY 77 VOIPECGGLGKAYVIDEESFMEYRYQIAEGCIRDLIEHP--SHSEK-----SSSY 129
DB 125 ATLPISTLGLBEAVYIIDESAFTLERLEIN-----SRFPQYFNTEKELLTSSSY 177
QY 130 Q--KQLOPERFLADYIERFICSYTBOIAVINMEKFLREHKDVRIVYIDSVTFHRRDPE 187
DB 178 HICRELTCEGLIQ-----RLSELEEII-----SKGVKLIYDSIASVVRKEPD 221
QY 188 -----DLAIFRPVYL-SGLSLKMTAKTYNLAIVLLNVTYTKF----- 224
DB 222 PKLGQIKERKNFGLGKASL-LKYLAGESFIVYILNTQTHLSGALPSQADLVSPADDL 280
QY 225 -----TEGSFQTLTAGDSWSHSCNTRLILHWNNGERNVAHL-DKSPSLPVASAPYAVTGK 278
DB 281 SLSEGSQSSGLVALAGNTWGHCVMTRLTILOYLDSERRQILIAKSPILAFTSFVYTIKE 340
QY 279 GIRDAVSSNHRK 290
DB 341 GL--VLOGHER 349
RESULT 11
ID DMCL_SOTBN STANDARD; PRT; 345 AA.
AC Q96449;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MEIOTIC RECOMBINATION PROTEIN DMCL HOMOLOG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]

RP SEQUENCE FROM N.A.
RA Hadi M.Z., Finer J.J.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PARTICIPATE IN METEOTIC RECOMBINATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U66836; AAB07025.1; -
DR HSSP; P03017; 2REB.
DR PROSITE; PS50162; RECA_2; 1.
DR PROSITE; PS50163; RECA_3; 1.
KW Meiosis; Cell cycle; ATP-binding; Nuclear protein.
FT NP BIND 134 141 ATP (POTENTIAL).
SQ SEQUENCE 345 AA; 37475 MW; 683CB96F923BAB6 CRC64;

Query Match 20.4%; Score 312; DB 1; Length 345;
Best Local Similarity 32.6%; Pred. No. 1.2e-19;
Matches 90; Conservative 46; Mismatches 108; Indels 32; Gaps 6;

OY 15 YVSGAQNAMDFSDLSQKHITTTSGDNDLLGGHCKEYTEIGVGVGKTLQIGLA 74
DB 93 YITGS---DALTKRKSVIRITGSGALDELGGVETSAITFAFGFRSGKTLAHTLC 148
OY 75 INVQIPVACGGLGKAVYIDTEGSMFWRVYQIAEGCRLDLEHPHSHKSSSVQKOLQ 134
DB 149 VSTOLPTMRGNGKVAYIDTEGTFPRDRIPIAE-----RCGMD 188
OY 135 PERFLADYIFRICSYTOIAVINMEKFLREHKDVRIIVSYTFHRODFE---DLAL 191
DB 189 PGAVLDNITVIRATYIEHQYMLGLAKMSE-EPFRLLIVDSYALFRVDFSGRGLAD 247
OY 192 RTRVLISGLSLKMKIATKYNAVLLNOVTKTEGSE---QLTLALGDSKSHSCNRLI 248
DB 248 RQOKLAQMLSLRIKIAEEFNVAAYMTNQYISDPGGGVFTDPKKRPGGHVLAHAATVRLM 307
OY 249 LHWNGNERVAHLDKSPSLPVASAPYATVGKIRDA 283
DB 308 FRKGGEORICKVEDAPMLPEAEAVFOITAGGIADA 343

RESULT 12
RADA_METUA STANDARD; PRT; 352 AA.
ID RADA_METUA
AC Q49593; Q58279;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA REPAIR PROTEIN RADA.
GN RADA OR M0869.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96251095; PubMed=8668545;
RA Sandler S.J., Satin L.H., Samra H.S., Clark A.J.;
RT "RecA-like genes from three archaean species with putative protein
RT products similar to Rad51 and Dmc1 proteins of the yeast
RT Saccharomyces cerevisiae."
RL Nucleic Acids Res. 24:2125-2132(1996).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=8680887;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierulage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Feldman J.F., Fuhrmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPAIR AND IN HOMOLOGOUS
CC RECOMBINATION AND REQUIRED FOR METEOTIC RECOMBINATION.
CC -1- SIMILARITY: STRONG TO EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U45312; AAC44122.1; -
DR EMBL; U67531; AAB98875.1; -
DR TIGR; M0868;
DR PROSITE; PS50162; RECA_2; 1.
DR PROSITE; PS50163; RECA_3; 1.
KW DNA damage; DNA recombination; ATP-binding; DNA-binding.
FT NP BIND 136 143 ATP (POTENTIAL).
SQ SEQUENCE 352 AA; 39084 MW; 861579C0C99AA29C CRC64;

Query Match 20.4%; Score 312; DB 1; Length 352;
Best Local Similarity 31.0%; Pred. No. 1.3e-19;
Matches 92; Conservative 47; Mismatches 106; Indels 52; Gaps 8;

OY 1 MDGSGSRNPOQKRYVSGAQNAMDFSDLSQKHITTTSGDNDLLGGHCKEYTEIG 60
DB 93 LGRKSGT---EVLQRKNIV-----KISTGSKNDLELGGGIESOSTEPFG 136
OY 61 VPGVGKTLQIGLAINVQIPVACGGLG-----KAVYIDTEGSMFWRVYQIAEGC 111
DB 137 MFGSGKTLQIAHQACVNLQCCERTVADATKDELINPKAVYIDTEGTFPRVQAAEYL 196
OY 112 IDLLEHPHSHKSSSVQKOLQPERFLADYIFRICSYTOIAVINMEKFLREHKDVR 171
DB 197 -----GLDGENEVLNIFVRAVNSDMQMLVAENVENLIREGHNK 236
OY 172 IYIIVSYTFHRODFE---EDLALTRVLSGLSLKMKIATKYNAVLLNOVTK--FTE 226
DB 237 LVIVDSLTSTFRTVEYIGRGALERQOKLGRHMAFLKMLADYINCVAIVNQAAARDALF 296
OY 227 GSFOLTLALGDSKSHSCNRLILH-WNGNERVAHLDKSPSLPVASAPYATVGKIR 282
DB 297 GPGEQALG-CHYGHATPFRIFLRKAKGKRVAKLVDSPHLPDAEMAFRTTEKGIND 352

RESULT 13
DLH1_CANAL STANDARD; PRT; 324 AA.
ID DLH1_CANAL
AC P50265;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE METEOTIC RECOMBINATION PROTEIN DLH1 (DMC1 HOMOLOG).
GN DLH1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Microsporite Saccharomycetales; Candida.

OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96363911; PubMed=8725225;
 RA Diemer A.C., Fink G.R.;
 RT "DLH1 is a functional Candida albicans homologue of the
 RL meiosis-specific gene DMCL1";
 RL Genetics 143:769-776(1996).
 CC -1- FUNCTION: REQUIRED FOR MEIOTIC RECOMBINATION, SYNAPTONEMAL COMPLEX
 CC FORMATION AND CELL CYCLE PROGRESSION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
 CC PROKARYOTIC RECA PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U39808; AAC49400.1; -
 DR PROSITE: PS50162; RECA_2; 1.
 DR PROSITE: PS50163; RECA_3; 1.
 KW Meiosis; Cell cycle; ATP-binding; Nuclear protein.
 FT NP_BIND 112 119 ATP (POTENTIAL).
 FT SEQUENCE 324 AA; 35281 MW; 95CD769514D9CF82 CRC64;
 SQ
 Query Match 19.9%; Score 304; DB 1; Length 324;
 Best Local Similarity 33.2%; Pred. No. 5,5e-19;
 Matches 87; Conservative 38; Mismatches 99; Indels 38; Gaps 7;
 OY 34 HTTSGDLDNDILGGGHCHEVTEIGVPGVKTOGLQIALNVOIPVCCGLGKAYI 93
 DB 86 HTTSGKQDEILGGGSGITGSGTVEGFRCKTQCHTLCAADLPDMGSGEGKAVI 145
 OY 94 DTGSGFMEVRYQIAE--GCIRDILHFRPHSHKSSSVQKOLPERLADIIYFRICSYT 151
 DB 146 DTGTFPRPRIRSIARVGVADIC-----LENISYARALNSE 183
 OY 152 EQIAVINMEKFLREHKDRIYIDSVTFHRODFE--DLALRTVLSGLSLKMKIAK 208
 DB 184 HOELVEOLGNELAE--GTERLLIVDSIMACFVDSGRLERKOOKHLSNLTVAE 242
 OY 209 TYNLAVLLNOVTTKFTGSGFQTLAL-----GDSMHSCTNRLILH-WNGNERYAHL 260
 DB 243 DYNIAVFLNVOYS--DPGASALFAAADGRKRVGGHVLAAHSATRLILKRGGERVAKL 300
 OY 261 DKSPLPVAAPYAVTGKGI 282
 DB 301 QDSPNMEKECVYVIGEGIKD 322
 RESULT 14
 RAS1_USTMA STANDARD; PRT; 339 AA.
 ID RAS1_USTMA
 AC Q99133;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE DNA REPAIR PROTEIN RAS1.
 GN RAS1.
 OS Ustilago maydis (Smut fungus).
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 CC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
 OX NCBI_TaxID=5270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97225207; PubMed=9071580;
 RA Ferguson D.O., Rice M.C., Rendi M.H., Kotani H., Kmiec E.B.,
 RA Holloman W.K.;

RT "Interaction between Ustilago maydis REC2 and RAD51 genes in DNA
 RT repair and mitotic recombination.";
 RL Genetics 145:243-251(1997).
 CC -1- FUNCTION: REQUIRED BOTH FOR RECOMBINATION AND FOR THE REPAIR OF
 CC DNA DAMAGE CAUSED BY X-RAYS (BY SIMILARITY).
 CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
 CC PROKARYOTIC RECA PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U62484; AAC61878.1; -
 DR InterPro: IPR000445; -
 DR Pfam: PF00633; HHH; 1.
 DR PROSITE: PS50162; RECA_2; 1.
 DR PROSITE: PS50163; RECA_3; 1.
 KW DNA damage; DNA repair; ATP-binding; DNA recombination.
 FT NP_BIND 127 134 ATP (POTENTIAL).
 FT SEQUENCE 339 AA; 36396 MW; DOE3676272FB0254 CRC64;
 SQ
 Query Match 19.6%; Score 299.5; DB 1; Length 339;
 Best Local Similarity 32.4%; Pred. No. 1.4e-18;
 Matches 84; Conservative 46; Mismatches 98; Indels 31; Gaps 6;
 OY 35 ITTSGDLDNDILGGGHCHEVTEIGVPGVKTOGLQIALNVOIPVCCGLGKAYI 94
 DB 102 ITTSGKNDLALGGGEMGTSTETLYGFRCKTQCHTLAVTCOLPDMGSGEGKCLYD 161
 OY 95 TEGSGFMEVRYQIAE--GCIRDILHFRPHSHKSSSVQKOLPERLADIIYFRICSYT 154
 DB 162 TETTFPRRLVAE-----RGLNGEEVLDVAARAVNADHQL 201
 OY 155 AVINMEKFLREHKDRIYIDSVTFHRODFE--DLALRTVLSGLSLKMKIAKTYN 211
 DB 202 QLMQASAMMAESR--FSLIYDLSLTSLYRDFSGRLSAROMHLAFLGLRLRDLDEF 260
 OY 212 LAVVLLNOVTTKFTGSGFQTLAL-----GDSMHSCTNRLILH-WNGNERYAHLKSPS 265
 DB 261 VAVVITNOVVAQ--VDGATAFADAKKPIGNTIVAHASTRLSLKRGNGRIGRIADSPC 319
 OY 266 LPVASAPYAVTGKIRAV 284
 DB 320 LPEADAVFAISPEGITDPV 338
 RESULT 15
 DMCL1_SCHPO STANDARD; PRT; 332 AA.
 ID DMCL1_SCHPO
 AC Q42634; Q42880;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE MEIOTIC RECOMBINATION PROTEIN DMCL1.
 GN DMCL1 OR DMP1 OR SPAC8B11.03C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
 OX Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A. AND FUNCTION.
 RX MEDLINE=20368631; PubMed=10908327;
 RA Fukushima K., Tanaka Y., Nabeshima K., Yoneki T., Tougan T.,
 RA Tanaka S., Nojima H.;
 RT "Dmcl1 of Schizosaccharomycetes pombe plays a role in meiotic
 RT recombination.";
 RL Nucleic Acids Res. 28:2709-2716(2000).

```

RN [2]
RP SEQUENCE FROM N.A.
RA Yamazaki A.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA McLean J., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: REQUIRED FOR MEIOTIC RECOMBINATION AND CELL CYCLE
CC PROGRESSION.
CC -! SUBCELLULAR LOCATION: NUCLEAR.
CC -! SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB008545; BAA28671.1; -
DR EMBL; D64035; BAA23984.1; -
DR EMBL; AL021817; CAAL7024.1; -
DR PROSITE; PS50162; RECA_2; 1.
DR PROSITE; PS50163; RECA_3; 1.
KW Meiosis; Cell cycle; ATP-binding; Nuclear protein.
FT NP_BIND 120 127 ATP (POTENTIAL).
FT CONFLICT 25 25 G -> GRLG (IN REF. 2).
FT CONFLICT 237 237 Q -> K (IN REF. 2).
SQ SEQUENCE 332 AA; 36408 MW; 86B73EAA491F7B02 CRC64;

```

```

Query Match 19.4%; Score 297.5; DB 1; Length 332;
Best Local Similarity 30.4%; Pred. No. 2.1e-18;
Matches 82; Conservative 46; Mismatches 103; Indels 39; Gaps 6;

OY 29 ELSQKH-----ITGSGDNDILGGHCKEVTETGIGVPGVGTQGLQIALINVOIPEC 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84 EISQNRKRYMSISQSEALNGLGGIOSMSITEYFGEFRCKTQMSHTLCVTAQLPRDM 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 84 GGLGKAVYIDTEGSEWERYVOIAEGCIRDILHFPHSHEKSSSVQKQLOPERFLADY 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 144 GGAEGKVAFIPTGCTFRPDRIKATAE-----RFGVDADQAMENIT 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 144 YFRCISYEQIAVINMKEFLREHKDVRIIVDSVTFHRODFE---DLALRTVLSGLS 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 VSRVYNSEQMEYITKLTIFAEDQYRLLVDSIMALFRVDSGRGELSERQOKLIML 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 201 LKIMKIATYINLAVYLNQVTTKFTGSGFQTLALGDS-----WSHSCNRLILH-WN 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 ARLNINISSEFNVAIVFTNQV---ADPGAAMMFASNDKRPVGVHMAHASATRLLLKGR 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 253 GNERYAHLDKSPSLPVASAPYAVTGKGIKD 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 GEERYAKLNDSPDMPEACSYVITPGGIAD 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: October 1, 2001, 16:29:24
 Job time: 117 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time 15.99 Seconds
(without alignments)
1400.583 Million cell updates/sec

Title: US-09-537-654-2

Perfect score: 1530
Sequence: 1 MDQSGSRNGPQOKTVSGAQ.....VTGKGRDAVSSNKRRAVT 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68:**
1: PIR1:**
2: PIR2:**
3: PIR3:**
4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	988	64.6	332	2	probable RAD51C-11
2	417	27.3	319	2	probable rada prot
3	387.5	25.3	324	2	DNA repair protein - Sul
4	354.5	23.2	337	2	DNA repair protein
5	352.5	23.0	311	2	DNA repair protein
6	345	22.5	356	2	DNA repair protein
7	337	22.0	334	2	meiosis-specific r
8	314.5	20.6	340	2	meiosis-specific r
9	313	20.5	340	2	meiosis-specific r
10	313	20.5	349	2	hypothetical 38.3K
11	312	20.4	345	2	RecA/Rad51/DMC1-11
12	312	20.4	352	2	DNA repair protein
13	309	20.2	357	2	rad51 protein, sho
14	309	20.2	391	2	hypothetical prote
15	304	19.9	324	2	DMC1/LIM15 homolog
16	298.5	19.5	336	2	meiotic recombina
17	297.5	19.4	332	2	DNA repair protein
18	293.5	19.2	386	2	DMC1/LIM15 homolog
19	290	19.0	342	2	JC4092
20	287.5	18.8	339	2	S35642
21	287.5	18.8	343	2	rad51 protein - Hal
22	281.5	18.4	339	2	rad51 protein homo
23	280.5	18.3	333	2	rad51 protein homo
24	280.5	18.3	353	2	recombination/rep
25	277.5	18.1	400	2	rad51 protein - ye
26	276	18.0	365	2	rad51 protein homo
27	272.5	17.8	342	2	rad51 protein homo
28	265.5	17.4	353	2	rad51 homolog mei-
29	260.5	17.0	354	2	DNA repair protein

30	255.5	16.7	529	2	probable DNA repai
31	242	15.8	353	2	probable RAD51B-11
32	241.5	15.8	460	2	rad51 protein - ye
33	215.5	14.1	220	2	DNA repair protein
34	207	13.5	236	2	DNA repair protein
35	206	13.5	221	1	DNA repair protein
36	200	13.1	239	2	recombinase PAR27
37	199	13.0	212	1	DNA repair protein
38	199	13.0	228	1	probable DNA repai
39	198.5	13.0	516	2	rad57 related prot
40	189	12.4	234	1	DNA repair protein
41	187	12.2	225	2	probable DNA repai
42	164.5	10.8	350	2	DNA repair protein
43	148	9.7	285	2	hypothetical prote
44	147.5	9.6	446	2	probable DNA repai
45	133	8.7	458	1	DNA repair protein

ALIGNMENTS

RESULT 1

F84888
probable RAD51C-like DNA repair protein [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84888
R:Lin, X.F. Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.I., Town, C.D., Fujii, C.Y.
M., Koo, H., Moffatt, K.S., Cronin, L.A., Shen, W., Vanden, S.E., Umayar, C.M., Venter
euss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M., Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MIMD:20083487
A:Accession: F84888
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <STO>
A:Cross-references: GB:AE002093; NID:g2583126; PIDN:AAB82635.1; GSPDB:GNO0139
C:Genetics:
A:Gene: At2g45280
A:Map position: 2

Query Match 64.6%; Score 988; DB 2; Length 332;

Best Local Similarity 66.5%; Pred. No. 2.9e-77;
Matches 183; Conservative 51; Mismatches 39; Indels 2; Gaps 2;

QY	19	AQNAAMFSDLSOKATTTTSGDLNDLGGGCHCKEYETIGVPGYKQGLQALINVO	78
DB	59	AKNAAMDMLHEEESLPRTTSCSDNLTGGGISCRTDTGVPGLKQIGLQISVNO	118
QY	79	IPVCGGLGKAVYIDPESFMYERYVQIAEGCIRDLIEHPHSHEKSSVQKQLOPERF	138
DB	119	IPRCGGLGKATYIDPESFMYERYVQIAEGCIRDLIEHPHSHEKSSVQKQLOPERF	178
QY	139	LADYPRICSYTEQALAVINMEKREKRDVRIYDSTVFHRQDEDLARTVLSG	198
DB	179	LENIFPRVCSYTEQALAVINMEKREKRDVRIYDSTVFHRQDEDLARTVLSG	237
QY	199	LSLKLMIATKYNTAVVLLNOVTKFEGSFQTLALGDSMSHSCYRRLILHNNGNERYA	258
DB	238	MAKFMKLAKFSLAVVLLNOVTKFEGSFQTLALGDSMSHSCYRRLILHNNGNERYA	297
QY	259	HLDKSPSLPVASAPYAVTGKGRDAVSSNKRRAVT	293
DB	298	YIDKSPSLPVASAPYAVTGKGRDAVSSNKRRAVT	331

RESULT 2

D72766
probable rada protein APE0119 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000


```

Oy      212 LAIVLLINQVTFEFTESFQLTAL-----GDSMHSCTNRLIILH-WNGENRYAHLEKS    263
        ||| | ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      255 VAVFLFNQVQS--DGCASLFFASADCKRRTIGGHVLAHASATRIILLRKRGDERVAKLQDS    312

Oy      264 PSLPVASAPYAVTGKIRDA    283
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      313 PMPEKECVYVIGEKGITDS    332

RESULT      8
JC4191
meiosis-specific recombination protein protein DMC1 homolog - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-Sep-1995 #sequence_revision 27-Oct-1995 #text_change 02-Feb-2001
C:Accession: Jc4191; S62353
R:Sato, S.; Kobayashi, T.; Hotta, Y.; Tabata, S.
DNA Res. 2, 147-150, 1995
A>Title: Characterization of a mouse recA-like gene specifically expressed in testis.
A:Reference number: JC4191; MUID:96038268
A:Accession: Jc4191
A:Molecule type: mRNA
A:Residues: 1-340 <SAT>
A:Cross-references: DDBJ:D58419; NID:9661475; PIDN:BAA09590.1; PID:9661476
A:Experimental source: testis
R:Habu, T.; Taki, T.; West, A.; Nishimune, Y.; Morita, T.
Nucleic Acids Res. 24, 470-477, 1996
A>Title: The mouse and human homologs of DMCI, the yeast meiosis-specific homologous recombination protein.
A:Reference number: S62353; MUID:96173646
A:Accession: S62353
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-340 <HAB>
A:Cross-references: EMBL:D64107; NID:9987666; PIDN:BAI0969.1; PID:91321647
C:Comment: This protein participates in the meiotic recombination process during mouse spermatogenesis.
A:Gene: MmLim15
C:Superfamily: yeast DNA repair protein RAD51
C:Keywords: nucleotide binding; P-loop; spermatogenesis; testis
F:126-133/Region: nucleotide-binding motif A (P-loop)
F:219-224/Region: nucleotide-binding motif B

Query Match          20.6%; Score 314.5; DB 2; Length 340;
Best Local Similarity 33.2%; Pred. No. 2.1e-19;
Matches 87; Conservative 43; Mismatches 97; Indels 35; Gaps 7;

Oy      34 HTSSSGDLNDLGSHCKEYTEIGVPVGKLTGLGIOLAINVOIPVECGLGSKANYI    93
        ||||| : : ||||| : | : | ||||| : | : | ||||| : | : | ||||| : | : |
Db     100 HTSSSQEDKLLGGISMAITFAFGFRGTQTLSHTCTAGLPPTGVSQGKIIFI    159

Oy      94 DTGGSFMERYVLQEGCIRDLIEHFPPSHSEKSSSVQKQLPERPLADIYFRICSYTEQ    153
        ||| : | : ||| : | : | ||| : | : | ||| : | : | ||| : | : | ||| : | : |
Db     160 DTGNFTRDR-----LRDLADFNVDH-----EAVDNVLYARAKVTSEHQ    199

Oy     154 IAVINYW-EKELREHKDVRIVIDSVTFERODEF--DLALRTVLGSLKIMKTIAKT    209
        : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     200 MELDIYVAAKFHEEAGIFKLLIIDISIMALFRYDFGREGELAEROCKLAQMLSRLQISEE    259

Oy     210 YNLAVLLNQVTKTBESFQTLIAL-----GDSMHSCTNRLIILH-WNGENRYAHL    261
        ||| : | : ||| : | : | ||| : | : | ||| : | : | ||| : | : | ||| : | : |
Db     260 YNVAAYEVINOMT--ADGCAITMTFOADPKRKPIGHIILAHASTTRISLRKRGGEELRIAKTY    316

Oy     262 KSPSLPVASAPYAVTGKIRDA    283
        || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     317 DSPKRPENEATPATTAIGSIGDA    338

RESULT      9
S62354
meiosis-specific recombination protein DMCI - human
N:Alternate names: meiosis-specific RecA-like protein; protein Lim15
C:Species: Homo sapiens (man)
```

C.Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 02-Feb-2001
C.Accession: S62354; J04333
R.Habu, T.; Takai, T.; West, A.; Nishimune, Y.; Morita, T.
Nucleic Acids Res. 24, 470-477, 1996
A.Title: The mouse and human homologs of DMCL1, the yeast meiosis-specific homologous
A.Reference number: S62353; MUID:96173646
A.Accession: S62354
A.Status: nucleic acid sequence not shown
A.Molecule type: mRNA
A.Residues: 1-340 <HAB>
A.Cross-References: EMBL:D64108; NID:g987656; PIDN:BA010970.1; PID:g1321636
A.Experimental source: testis
R.Sato, S.; Seki, N.; Hotta, Y.; Tabata, S.
DNA Res. 2, 183-186, 1995
A.Title: Expression profiles of a human gene identified as a structural homologue of
A.Reference number: J04333; MUID:96127532
A.Accession: J04333
A.Molecule type: mRNA
A.Residues: 1-36, 'T', '38-182, 'P', 184-340 <SAT>
A.Cross-References: DDBJ:D63882; NID:g1066000; PIDN:BA009932.1; PID:g1066001
A.Experimental source: testis
C.Comment: This protein plays an essential role in meiotic recombination.
C.Genetics:
A.Gene: hsc01
C.Superfamily: yeast DNA repair protein RAD51
C.Keywords: DNA binding; meiosis; nucleotide binding; P-loop; testis
F.126-133/Region: nucleotide-binding motif A (P-loop)
F.219-223/Region: nucleotide-binding motif B

[illegible]

```

Query Match          20.5%; Score 313; DB 2; Length 349;
Best Local Similarity 33.3%; Pred. No. 2.9e-19;
Matches 92; Conservative 48; Mismatches 104; Indels 32; Gaps 7

15 YVSGAQNAMDMSDELSOKHITTTGSDLLDILGGGHNCKEVEIGGVPVGKQTOLGIOLA 74
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 97 YITGS----DVLKRRKSVIRITTSQALDELLGGGIEFTQITEARCEPFGSGTQIAHTLC 152
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

75 INVQIPVBCGGLGCAVAVYIDTEGSPMERVRYOIAEGCINDILLENHPHSHKSSSVOKOLO 134
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 153 VSTQLPVSMHGNGKGVAVYIDTEGTRPRDRIVPIAEFGMD-----ASAV----- 196
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

135 PERFLADLYFFRCGYTBDIAVINMKEFLREHKDVRIVYIDSVYTFHFRODE--DLAL 191
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 197 ----IDNIIYAAYVYEHQYNLLALAKMS-EPFRLLIVSVIALFVDFSGRGELAE 251
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

192 RTRVLGSLKLMKTAQTYNLAVALVLTQVTKFTTEGSGF--QLTIALGDSWHSCTNRLI 248
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 252 RQOKLAQMLSLRLTKIAEEFNNAVYMTNVOYIAPGGGMFTSDPKKRRGGHYLAHAATVRLM 311
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

249 LH-WNGNERRYAHLDKSPSLPVASADYAVTGGKIRDA 283
   |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 312 LKKGGEQRVCKIFDAPNLPESAFAVQTIPGGVADA 347
   |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

RESULT 11
T08838
Reca/Rad51/DMC1-like protein - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
C:Accession: T08838
R:Hadji, M.Z.; Finer, J.J.
submitted to the EMBL Data Library, August 1996
A:Description: RecA like gene from soybean.
A:Reference number: Z16485
A:Accession: T08838
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-345 <HAD>
A:Cross-references: EMBL:U66836; NID:g1518156; PID:g1518157
C:Function:
A:Description: recombinase
C:Superfamily: yeast DNA repair protein RAD51

Query Match          20.4%; Score 312; DB 2; Length 345;
Best Local Similarity 32.6%; Pred. No. 3.4e-19;
Matches 90; Conservative 46; Mismatches 109; Indels 32; Gaps 6;

15 YVSGAQNAMDMSDELSOKHITTTGSDLLDILGGGHNCKEVEIGGVPVGKQTOLGIOLA 74
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 93 YITGS----DALTKRKSIVIRITTSQALDELLGGGVEISATIEARCEPFGSGTQIAHTLC 148
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

75 INVQIPVBCGGLGCAVAVYIDTEGSPMERVRYOIAEGCINDILLENHPHSHKSSSVOKOLO 134
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 149 VSTQLPVSMHGNGKGVAVYIDTEGTRPRDRIVPIAE-----RFGMD----- 188
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

135 PERFLADLYFFRCGYTBDIAVINMKEFLREHKDVRIVYIDSVYTFHFRODE--DLAL 191
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 189 PGAVLDNIIYARAATYIEHQYNLLGLAKMS-EPFRLLIVDSVIALFVDFSGRGELAD 247
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

192 RTRVLGSLKLMKTAQTYNLAVALVLTQVTKFTTEGSGF--QLTIALGDSWHSCTNRLI 248
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 248 RQOKLAQMLSLRLTKIAEEFNNAVYMTNVOYIAPGGGVEVTDPKKRRGGHYLAHAATVRLM 307
   :::: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

249 LH-WNGNERRYAHLDKSPSLPVASADYAVTGGKIRDA 283
   |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db 308 FRKGGEQRICKVFDAPNLPESAFAVQTIPAGGIADA 343
   |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |

RESULT 12
E64408
DNA repair protein rada - Methanococcus jannaschii

```

C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: E64408; S71095; S71096
R:Bult, C.U.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
e, Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: E64408
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-352 <BU>
A:Cross-references: GB:067531; GB:L77117; NID:q2826342; PIDN:AA98875.1; PID:91591553
R:Sandler, S.J.; Satlin, L.H.; Samra, H.S.; Clark, A.J.
Nucleic Acids Res. 24, 2125-2132, 1996
A:Title: recA-like genes from three archaean species with putative protein products s
A:Reference number: S71093; MUID:96251095
A:Accession: S71095
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 33-352 <SA>
A:Cross-references: EMBL:U45312
C:Genetics:
A:Gene: rada
A:Map position: FOR791863-792921
C:Superfamily: yeast DNA repair protein RAD51

Query March 20.4% Score 312; DB 2; Length 352;
Best Local Similarity 31.0%; Pred. No. 3.5e-19;
Matches 92; Conservative 47; Mismatches 106; Indels 52; Gaps 8;

QY 1 MGDSGSHNGPQQKYVSGAONAMDMFSDLSQKHITTGSGDLNDLGGCINCKEVEITEIG 60
:
Db LGFSKGT-----EVLSQRKNIM-----KLSTGSKNLDEILGGLESQSVTEPAG 136
:
QY 61 VPGKQTQLGLQIALINVOIPECCGLG-----KAYIIDEGSSMVARVQIQAECC 111
:
Db MFSGSKQTQIIQAACVNLCQCPERIVADDAIKDELINEPAAVVIDEGTFRPRIVQMAEAL 196
:
QY 112 IRDILEHPHSHKSSSQOKLOPERFLADIYPYRISCYTOIAVINMEKEFLREHDH 171
:
Db 197 -----GLDGNEVLNNIFVARAINSDMQMLXAENVLMIREGNIRK 236
:
QY 172 IVIDSVTFEHRRDF--EDLALRTRVLSGLSLMKIAKTYYNALVLLNOVTK--FTE 226
:
Db 237 LVIVDSLSTFTERTETYGKGLAEARQOKIGRMHATLNKLADIYNCVVIVTQVAAPDALF 296
:
QY 227 GSPQLTALGDSWHSCHNRLIH-WNCGNEKXATLDKSPSPVASAPYAVYGRKIRD 262
:
Db 297 GPSEQATG-GHIVGHAAFTFLFLKKAKGDKKRAVLKDYSPHLDPDAEMAFRIETKGIDH 352

RESULT 13
T37305
radsl protein, short isoform - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T37305; T43059
C:Takamami, T.; Sato, S.; Ishihara, T.; Katsura, I.; Takahashi, H.; Higashitani, A.
DNA Res. 5, 373-377, 1998
A>Title: Characterization of a Caenorhabditis elegans recA-like gene Ce-rdh-1 involve
A:Reference number: Z21672; MUID:99156232
A:Accession: T37305
A>Status: preliminary; translated from GB/EMBL/DDBB
A:Molecule type: mRNA
A:Residues: 1-357 <TRAK>
A:Cross-references: EMBL:AB011382; NID:g2913896; PIDN:BAZ482.1; PID:g2913897
Mol. Genet. 260, 289-294, 1998
R:Rinaldo, C.; Eberle, S.; Rocco, V.; La Volpe, A.
Title: The Caenorhabditis elegans RAD51 homolog is transcribed into two alternative

Tue Oct 2 11:16:28 2001

us-09-537-654-2.rpr

Page 7

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 05:00:54 ; Search time 124.96 seconds
(without alignments)
7406.586 Million cell updates/sec

Title: US-09-537-654-1

Perfect score: 1474
Sequence: 1 tgcaccacagctccgcact.....aaaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn/NA1980.DAT.*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn/NA1981.DAT.*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn/NA1982.DAT.*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn/NA1983.DAT.*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn/NA1984.DAT.*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn/NA1985.DAT.*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn/NA1986.DAT.*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn/NA1987.DAT.*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn/NA1988.DAT.*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn/NA1989.DAT.*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn/NA1990.DAT.*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn/NA1991.DAT.*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn/NA1992.DAT.*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn/NA1993.DAT.*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn/NA1994.DAT.*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn/NA1995.DAT.*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn/NA1996.DAT.*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn/NA1997.DAT.*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn/NA1998.DAT.*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn/NA1999.DAT.*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn/NA2000.DAT.*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1474	100.0	1474	22	MAIZE RAD51 orthol
2	1305.4	88.6	1333	22	MAIZE RAD51 orthol
3	1243.2	84.3	1459	22	MAIZE RAD51 orthol
4	178.4	12.1	1325	21	HTRM clone 2398682
5	90.2	6.1	936	22	OLIGONUCLEOTIDE D1
6	90.2	6.1	936	22	OLIGONUCLEOTIDE D1
7	90.2	6.1	936	22	OLIGONUCLEOTIDE D1
8	90.2	6.1	936	22	OLIGONUCLEOTIDE D2
9	90.2	6.1	936	22	OLIGONUCLEOTIDE D2
10	89.8	6.1	936	22	OLIGONUCLEOTIDE D1
11	89.8	6.1	936	22	OLIGONUCLEOTIDE D1

12	89.8	6.1	936	22	AAF58254	OLIGONUCLEOTIDE D1
13	89.8	6.1	936	22	AAF58257	OLIGONUCLEOTIDE D1
14	89.8	6.1	936	22	AAF58259	OLIGONUCLEOTIDE D2
15	89.8	6.1	936	22	AAF58262	OLIGONUCLEOTIDE D2
16	89.8	6.1	938	22	AAF58255	OLIGONUCLEOTIDE D1
17	65.8	4.5	1020	20	AAZ11866	ZMRAD51A nucleotid
18	65.8	4.5	1568	22	AAZ11868	ZMRAD51A cDNA clon
19	65	4.4	1585	22	AAZ02318	Trilicium aestivum
20	64.4	4.4	1020	20	AAZ11870	ZMRAD51B nucleotid
21	64.4	4.4	1574	20	AAZ11872	ZMRAD51B cDNA clon
22	62.8	4.3	1219	22	AAZ02316	Glycine max RAD51
23	57.8	3.9	1682	16	AAQ91763	Human Rad51 gene,
24	57.8	3.9	2229	21	AAA37710	Human Rad51 mRNA.
25	55.8	3.8	1362	21	AAZ47934	Arabidopsis thalia
26	55.8	3.8	1408	16	AAQ91761	Mouse Rad51 gene,
27	55.8	3.8	1755	15	AAQ64088	Mouse gene particl
28	53.8	3.6	3776	17	AAZ42903	TRP-1 protein codi
29	53.8	3.6	3994	21	AAZ6475	Human OREF2030
30	52.8	3.6	2639	21	AAZ58108	Human PRO1788 nucl
31	52.8	3.6	2639	21	AAA37137	Human PRO1788 (UNQ
32	52.8	3.6	2639	22	AAZ54480	Probe #56 used in
33	52.4	3.6	2422	21	AAZ64112	Human OREF ORF1667
34	52.2	3.5	887	21	AAZ78089	Human cancer assoc
35	52	3.5	2737	21	AAZ64478	Human H38087 (clon
36	51	3.5	3447	21	AAZ76035	Human OREF ORF1590
37	50.8	3.4	876	21	AAZ9201	Human secreted pro
38	49.8	3.4	1522	21	AAZ93480	Human secreted pro
39	49.8	3.4	1641	17	AAZ15758	Human interleukin-
40	49.8	3.4	4091	11	AAZ004035	Stem cell leukemia
41	49.8	3.4	4199	11	AAZ005330	Stem cell leukemia
42	49.2	3.3	1483	21	AAZ65060	Membrane-bound pro
43	49.2	3.3	1484	22	AAZ44206	Human PRO1105 (UNQ
44	48.6	3.3	443	21	AAZ75749	cDNA encoding a de
45	48.6	3.3	553	21	AAZ79990	Human secreted pro

ALIGNMENTS

RESULT 1	
AAZ02118	
ID	AAZ02118 standard; cDNA; 1474 BP.
AC	AAZ02118:
XX	
DT	28-MAR-2001 (first entry)
XX	
DE	Maize RAD51 orthologue #1 cDNA.
XX	
KW	Maize; RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;
KM	homologous recombination; transformation; gene targeting;
KW	transgenic plant; DNA repair; ss.
OS	
XX	
OS	Zea mays.
XX	
FX	
FX	Key
FT	CDS
FT	Location/Qualifiers
FT	310..1194
FT	/*tag= a
FT	/product= "Maize RAD51 orthologue protein #1"
PN	W0200068370-A2.
XX	
PD	16-NOV-2000.
XX	
PF	05-APR-2000; 2000MO-US09010.
XX	
PR	05-MAY-1999; 99US-0132582.
XX	
PA	(PION-) PIONEER HI-BRED INT INC.
XX	
PI	Mahajan PB, Shi J;
XX	
DR	WPI; 2001-016092/02.

DR P-PSDB; AAY1974.

XX New maize RAD51 polynucleotides, useful for e.g. controlling
PT recombination or transformation efficiency in plants, or as probes or
PT amplification primers for detecting, quantifying or isolating gene
PT transcripts

XX Claim 1; Page 68-69; 76pp; English.

XX The present sequence is a maize RAD51 orthologue #1 cDNA. RAD51 is one
CC of the genes of the RAD52 epistasis group that is involved in double
CC strand breaks (DSBs) repair by homologous recombination. Control of
CC homologous recombination by modulating RAD51 is used for improving
CC transformation and gene targeting in transgenic plants. The RAD51
CC nucleotide may be used as probes or amplification primers for detecting,
CC quantifying or isolating gene transcripts, in detecting deficiencies in
CC the mRNA level during screening for desired transgenic plants, for
CC detecting gene mutations, for monitoring upregulation of expression or
CC changes in enzyme activity, for detecting any number of allelic variants,
CC orthologues or paralogues of the gene, or for site directed mutagenesis
CC in eukaryotic cells. It may also be used for recombinant expression of
CC its encoded polypeptide, or for use as immunogen in preparing and/or
CC screening of antibodies, and in sense or antisense suppression of one or
CC more genes in a host cell, tissue or plant. The RAD51 proteins may be
CC used in assays to agonise or antagonise the enzyme function, or as
CC immunogens or antigens to obtain antibodies.

XX Sequence 1474 BP; 415 A; 337 C; 369 G; 353 T; 0 other;

Query Match 100.0%; Score 1474; DB 22; Length 1474;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgaccacgcgtccgactgactgactccacatgctgcagattcgctgtccc 60
DB 1 tcgaccacgcgtccgactgactgactccacatgctgcagattcgctgtccc 60
QY 61 cggagcccaaaagcggcggtgagccgagcccgagagacgagcgcgagactccc 120
DB 61 cggagcccaaaagcggcggtgagccgagcccgagagacgagcgcgagactccc 120
QY 121 cctaagcagacagcgcggtgagcgtgagcgtgctggtgagccagcgagagctac 180
DB 121 cctaagcagacagcgcggtgagcgtgagcgtgctggtgagccagcgagagctac 180
QY 181 gaggcggttgaggagacagatgagaggtgaggtgagctacacggtcgaggtctgca 240
DB 181 gaggcggttgaggagacagatgagaggtgaggtgagctacacggtcgaggtctgca 240
QY 241 gatctggaatccgcacgtctctctctccccaatcagttaccactccaagt 300
DB 241 gatctggaatccgcacgtctctctctccccaatcagttaccactccaagt 300
QY 301 gcaatcacatgagagatcaatctgctctagaatgagacacagaaagttagttca 360
DB 301 gcaatcacatgagagatcaatctgctctagaatgagacacagaaagttagttca 360
QY 361 ggaagcccaaatgctgagatgtctctctgtagtgcgtgacagaaacacatcact 420
DB 361 ggaagcccaaatgctgagatgtctctctgtagtgcgtgacagaaacacatcact 420
QY 421 ggtctgtgacctaaagacatctgtgtggtgaggtacactgacaaagaattactag 480
DB 421 ggtctgtgacctaaagacatctgtgtggtgaggtacactgacaaagaattactag 480
QY 481 atcgtgtgctccaggggttgtaaactcaactgggagttcaactgacataatgta 540
DB 481 atcgtgtgctccaggggttgtaaactcaactgggagttcaactgacataatgta 540
QY 541 caaatcccaagtgaatgtgtgtgctgtgtggaagacgtttatatagatcacagagcc 600
DB 541 caaatcccaagtgaatgtgtgtgctgtgtggaagacgtttatatagatcacagagcc 600

QY 601 agttcatggttgaacgtgtctaccagattgtgaaggtgtattagggacatctggag 660
DB 601 agttcatggttgaacgtgtctaccagattgtgaaggtgtattagggacatctggag 660
QY 661 cacttccgcacagacatgagaagtcctctctcttctcccaaaacatcacagctgagcgt 720
DB 661 cacttccgcacagacatgagaagtcctctctcttctcccaaaacatcacagctgagcgt 720
QY 721 ttctctgaggtatctattacttccggatatagtacacccgaacaaattgcagtcata 780
DB 721 ttctctgaggtatctattacttccggatatagtacacccgaacaaattgcagtcata 780
QY 781 aactacatgaggaaggtccctcagagacataaagaatgtgcgtatagttattgttagt 840
DB 781 aactacatgaggaaggtccctcagagacataaagaatgtgcgtatagttattgttagt 840
QY 841 gttacttccacttgcacaaagatttgaagatctgacactgagagaccagagtgtcaat 900
DB 841 gttacttccacttgcacaaagatttgaagatctgacactgagagaccagagtgtcaat 900
QY 901 ggtattatcattgaagttgaagattgcaagacacataacttggcagttgtctgtgtg 960
DB 901 ggtattatcattgaagttgaagattgcaagacacataacttggcagttgtctgtgtg 960
QY 961 aaccaagtcactaataattacagaaggtcatttcaattgactcttgcctaggtgac 1020
DB 961 aaccaagtcactaataattacagaaggtcatttcaattgactcttgcctaggtgac 1020
QY 1021 agctgttcccaactatgacagacacggttgaattctgacatggaatgggaacagatlac 1080
DB 1021 agctgttcccaactatgacagacacggttgaattctgacatggaatgggaacagatlac 1080
QY 1081 gcaacatctgataaggtctccttccacttccagtagtccctgacacccgataatgagtgacagcc 1140
DB 1081 gcaacatctgataaggtctccttccacttccagtagtccctgacacccgataatgagtgacagcc 1140
QY 1141 aaagggatgagatgctgtgaggtcacaacacacaaagcgagcccgagtaacgtacattct 1200
DB 1141 aaagggatgagatgctgtgaggtcacaacacacaaagcgagcccgagtaacgtacattct 1200
QY 1201 tgggtgtaagcactgtgatttccactacgctctctctgacgcttcttccgacatgattctt 1260
DB 1201 tgggtgtaagcactgtgatttccactacgctctctctgacgcttcttccgacatgattctt 1260
QY 1261 ggaactgaggtgagagctgagagatgataccatttgtgtgattctccagttgcttctg 1320
DB 1261 ggaactgaggtgagagctgagagatgataccatttgtgtgattctccagttgcttctg 1320
QY 1321 cgttgcctacccaacacacttaagagagaagtaatacaacagacagctaatatagtg 1380
DB 1321 cgttgcctacccaacacacttaagagagaagtaatacaacagacagctaatatagtg 1380
QY 1381 ttgtgtatcgaacatctgcccactgacatcagtaagaagcctataatagcgggcaaaa 1440
DB 1381 ttgtgtatcgaacatctgcccactgacatcagtaagaagcctataatagcgggcaaaa 1440
QY 1441 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474
DB 1441 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1474

RESULT 2

AAD02120 standard; cDNA; 1333 BP.

XX AAD02120;

XX 28-MAR-2001 (first entry)

XX Maize RAD51 orthologue #3 cDNA.

KW Maize; RAD51 orthologue; RAD52 epistasis group; double strand break; DSB;

KW homologous recombination; transformation; gene targeting;
transgenic plant; DNA repair; ss.

XX Zea mays.

XX Key Location/Qualifiers
XX FT 169..1053
XX CDS /*tag= a
XX /product= "Maize RAD51 orthologue protein #3"

XX W0200068370-A2.

XX 16-NOV-2000.

XX 05-APR-2000; 2000MO-US09010.

XX 05-MAY-1999; 99US-0132582.

XX (PION-) PIONEER HI-BRED INT. INC.

XX Mahajan PB, Shi J.

XX WPI; 2001-016092/02.

XX P-PSDB; AAY1976.

PT New maize RAD51 polynucleotides, useful for e.g. controlling
PT recombination or transformation efficiency in plants, or as probes or
PT amplification primers for detecting, quantifying or isolating gene
PT transcripts.

XX Claim 1; Page 72-74; 76pp; English.

CC The present sequence is a maize RAD51 orthologue #3 cDNA. RAD51 is one
CC of the genes of the RAD52 epistasis group that is involved in double
CC strand breaks (DSBs) repair by homologous recombination. Control of
CC homologous recombination by modulating RAD51 is used for improving
CC transformation and gene targeting in transgenic plants. The RAD51
CC nucleotide may be used as probes or amplification primers for detecting,
CC quantifying or isolating gene transcripts, in detecting deficiencies in
CC the mRNA level during screening for desired transgenic plants, for
CC detecting gene mutations, for monitoring upregulation of expression or
CC changes in enzyme activity, for detecting any number of allelic variants,
CC orthologues or paralogues of the gene, or for site directed mutagenesis
CC in eukaryotic cells. It may also be used for recombinant expression of
CC its encoded polypeptide, or for use as immunogen in preparing and/or
CC screening of antibodies, and in sense or antisense suppression of one or
CC more genes in a host cell, tissue or plant. The RAD51 proteins may be
CC used in assays to agonise or antagonise the enzyme function, or as
CC immunogens or antigens to obtain antibodies.

XX Sequence 1333 BP; 377 A; 284 C; 326 G; 345 T; 1 other;

Query Match 88.6%; Score 1305.4; DB 22; Length 1333;

Best Local Similarity 98.6%; Pred. No. 0;

Matches 1315; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 142 cgaagtaagcgcgtgcgtgcgcacccagagagcgttctgtgagagcagata 201
DB 1 cgaagtaagcgcgtgcgtgcgcacccagagagcgttctgtgagagcagata 60
QY 202 tgaagagtgagagtggtacacacggtgcgcgtctgagataactaaatccacatgca 261
DB 61 tgaagagtgagagtggtacacacggtgcgcgtctgagataactaaatccacatgca 120
QY 262 gttctctcttcccccaatcaagtaacaccccaagtggaatccatggagatcaaa 321
DB 121 gttctctcttcccccaatcaagtaacaccccaagtggaatccatggagatcaaa 180
QY 322 tctgctctgaagagcacaacagagtagcttccagagccagaaatgctctggat 381
DB 181 tctgctctgaagagcacaacagagtagcttccagagccagaaatgctctggat 240

QY 382 atgtctctgaagagctgtccacagaaacacatcactactgttctgtgacatcaatgac 441
DB 241 atgtctctgaagagctgtccacagaaacacatcactactgttctgtgacatcaatgac 300
QY 442 atactgtgtgcggagatcactgtcaagaagatctactgagatcgtgtcccaagggtc 501
DB 301 atactgtgtgcggagatcactgtcaagaagatctactgagatcgtgtcccaagggtc 360
QY 502 ggtaaacccaactgtggagttcaactagcaatcaatgtaacaaatcccgatggatgtgt 561
DB 361 ggtaaacccaactgtggagttcaactagcaatcaatgtaacaaatcccgatggatgtgt 420
QY 562 ggcctgtgtggagaaagcagttatataagatacagagggcagttcatgttgaacgttc 621
DB 421 ggcctgtgtggagaaagcagttatataagatacagagggcagttcatgttgaacgttc 480
QY 622 taccagatgtgtgaagagtgatattagggacatactgtgaacacttcgcagacagatgag 681
DB 481 taccagatgtgtgaagagtgatattagggacatactgtgaacacttcgcagacagatgag 540
QY 682 aagctctctgtcccaaaaacaattacagcgtgtgagcttctcgtggagatatactatc 741
DB 541 aagctctctgtcccaaaaacaattacagcgtgtgagcttctcgtggagatatactatc 600
QY 742 ttcggatatagtcagttacacccgacaacaaattgtagtcataactatgagaagttcttc 801
DB 601 ttcggatatagtcagttacacccgacaacaaattgtagtcataactatgagaagttcttc 660
QY 802 agagagcataaagatgtgcgtatagttatattatgataagtgtaacttccacttcgaca 861
DB 661 agagagcataaagatgtgcgtatagttatattatgataagtgtaacttccacttcgaca 720
QY 862 gattttgaagatctgtgcctctgagccagagagtgctcaagtgtatctattgaagtta 921
DB 721 gattttgaagatctgtgcctctgagccagagagtgctcaagtgtatctattgaagtta 780
QY 922 aagatgtcaaaaacacataactgtgcagttgtctgtgtgaacaaagtaactacataat 981
DB 781 aagatgtcaaaaacacataactgtgcagttgtctgtgtgaacaaagtaactacataat 840
QY 982 acagaaaggtcatttcaattgactctgtctagttgacagctgtgtccactatgacg 1041
DB 841 acagaaaggtcatttcaattgactctgtctagttgacagctgtgtccactatgacg 900
QY 1042 aaccggttatcttcacgtcagatgtggaacggaacatacgcacattgtataagttct 1101
DB 901 aaccggttatcttcacgtcagatgtggaacggaacatacgcacattgtataagttct 960
QY 1102 tcaattccagtagctcagcacccglatgtcagtgacaggaacgaagattagatgtgtg 1161
DB 961 tcaattccagtagctcagcacccglatgtcagtgacaggaacgaagattagatgtgtg 1020
QY 1162 agctcaaacacacagcagcccgagtaacgtatcattgtgtcaagcattgtatgt 1221
DB 1021 agctcaaacacacagcagcccgagtaacgtatcattgtgtcaagcattgtatgt 1080
QY 1222 ccactacgctctgcagcttcttcgcacatgatacttttgaacagtgaggtgaacgtg 1281
DB 1081 ccactacgctctgcagcttcttcgcacatgatacttttgaacagtgaggtgaacgtg 1140
QY 1282 agaatagtaacatttgtgtatctcagttgtctgtgtgcgttgaacacacacact 1341
DB 1141 agaatagtaacatttgtgtatctcagttgtctgtgtgtgcgttgaacacacacact 1200
QY 1342 aagagagaagtaataacacacagacagagcgtatataagtggtttgtaactgtgc 1401
DB 1201 aagagagaagtaataacacacagacagagcgtatataagtggtttgtaactgtgc 1260
QY 1402 ccattgtacattcagtaagcctataatagcggcaaaaaaataaaaaaataaaaaa 1461
DB 1261 ccattgtacattcagtaagcctataatagcggcgtatataatagtgcttctcgtatcaaaa 1320
QY 1462 aaaaaaataaaaaa 1474

[illegible]

[illegible]


```

PI Unek RM;
XX
XX DR WPI: 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface _
XX
XX
XX Example 6; Page 128; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 6.1%; Score 89.8; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 1.3e-13;
Matches 7; Conservative 454; Mismatches 316; Indels 0; Gaps 0

QY 698 aaaaacaattacagcctgagcgtttctcctggcgagatcttacttcgcgatatgcagtt 757
DB 9 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 68
QY 758 acacgaacaacattgcagtcataactatactatgagagaagttccacgagagcataaagatg 817
DB 69 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 128
QY 818 tgcgtatgattatattgatgatgtgttacttccacttcgcacaagaatttgaagatctgg 877
DB 129 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188
QY 878 cactgaggaacagagtcgaagtgatgattatcatctgaagttaagttaagattgcaaacat 937
DB 189 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248
QY 938 ataactggcagttgctctgttgaaccaaagtcactactaattacagaaggtcatcttc 997
DB 249 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 308
QY 998 aattgactcttgctcaggttgacagcgtgtccactcatgcagcaacogtltgattctgc 1057
DB 309 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368
QY 1058 actgaatggaagcaacagatagacacatctgataagctctcctcacttcagtagcct 1117
DB 369 www.wwwwwwwwwwwwwwwwwwwwwwwwgcttcttgaaggggggggggggggggggggg 428
QY 1118 cagcaacggtatgcattgacagcgaagagatagagatgctgtgagctcaaaccaagc 1177
DB 429 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488
QY 1178 gaggccgagtaacgtagacattctgtgtcgaagcactgtatgtccacttaagctctgca 1237
DB 489 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 548
QY 1238 gcttcttcgcacatgatacttttgactagtgaggtgagactgagaaatgatacat 1297
DB 549 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 608
QY 1298 gttgattctcagttgcttggcogtltgctaccacaacacttaagaagagataa 1357
DB 609 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 668
QY 1358 caacagaaagcgtaataatagtgtttgcatacgaactcggccatcgtatcattcagtt 1417
DB 669 www.wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 728
QY 1418 aaagcctataatagcgggcaaaaaaanaaaaaaanaaaaaaanaaaaaa 1474

```

Db 729 785

RESULT 15

AAF58262 standard; DNA; 936 BP.

AAF58262;

24-APR-2001 (first entry)

Oligonucleotide D2007.

Electron-transfer group; ETM; mismatch; genotyping;

Synthetic.

WO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000MO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umea RM;

WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in

hybridization assays, e.g. for genotyping, allowing repeat analyses on

a single surface

Example 6; Page 128; 159pp; English.

The present invention relates to a composition comprising two nucleic

acids each containing an electron-transfer group (ETM) having

different redox potentials. The invention is used for electronic

detection of nucleic acids, especially of substitutions (mismatches)

and single-nucleotide polymorphisms, e.g. for genotyping,

monitoring gene expression.

Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 6.1%; Score 89.8; DB 22; Length 936;

Best Local Similarity 0.9%; Pred. NO. 1.3e-13;

Matches 7; Conservative 454; Mismatches 316; Indels 0; Gaps 0;

698 aaaaacattacagcgtctctcgtcgagatctatctacgtatcgatg 757

9 68

758 aacacgacaaatgacgtacataactacatgagaagttcctcagagacataaagatg 817

69 128

818 tgcgtatagttatgtagtcttacttccacttcgacaaagatttgaagatcgg 877

129 188

878 cactgagaccagagtgctagtgattatcattgaagtaagaaatgcaagacat 937

189 248

938 ataactggcagtgctctgttgaaccaagtcactactaaattacagaaggtcattc 997

249 308

Qy 998 aattgactcttgctctagtgagacagctgtccactactgacgaaccggtgattcgc 1057

309 368

Qy 1058 actggaatggagaacgaacatacgacacttgataagctctcctcactccagtagcct 1117

369 428

Qy 1118 cagcaccgtatgacgtacaggaagattagagatgctgtgacgtccaaaccaagc 1177

429 488

Qy 1178 gagccagatlaacgtacgtctcttggtcaagcactgtatgtccactacgtcctgca 1237

489 548

Qy 1238 gcttcttcgcacatgacatcttcttgactagtgagtgactgagaaatgacattt 1297

549 608

Qy 1298 gttagattcagttgcttctgtgcgttgctacccaacacttaagagagaataata 1357

609 668

Qy 1358 caacagacaggtacataatagtggtttatcttgacacatctgcccacgtacattcagt 1417

669 728

Qy 1418 aagcctataatagcggcgaacaaacaaacaaacaaacaaacaaacaaacaaacaa 1474

729 785

Search completed: October 2, 2001, 05:51.15

Job time: 3021 sec

Tue Oct 2 11:16:24 2001

us-09-537-654-1.rng

Page 14

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time 12.27 Seconds

(Without alignments)
493.363 Million cell updates/sec

Title: US-09-537-654-2

Perfect score: 1530
1 MODGSGNRNGPOQKRVSGAQ.....VTGKGRDAVSSNKKRAVRT 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	21.1	350	4	US-09-157-603-1
2	323	21.1	350	4	US-09-587-436-1
3	313	20.5	350	4	US-09-157-603-3
4	313	20.5	350	4	US-09-587-436-3
5	281.5	18.4	339	3	US-08-758-280-2
6	281.5	18.4	339	3	US-08-664-614A-2
7	280.5	18.3	339	3	US-08-758-280-1
8	280.5	18.3	339	3	US-08-964-614A-1
9	126	8.2	781	2	US-08-373-134D-2
10	126	8.2	781	2	US-09-114-637-2
11	107	7.0	451	3	US-08-969-644-8
12	107	7.0	451	3	US-08-444-189-8
13	107	7.0	451	3	US-08-444-189-8
14	101	6.6	273	1	US-08-215-828A-2
15	101	6.6	273	1	US-08-446-920-2
16	101	6.6	347	1	US-08-446-920-10
17	96	6.3	388	1	US-08-261-110A-4
18	96	6.3	438	1	US-08-261-110A-2
19	93.5	6.1	340	1	US-08-097-831-2
20	91.5	6.0	429	2	US-08-310-912A-109
21	91.5	6.0	429	5	PCT-US95-04589-109
22	91.5	6.0	1824	4	US-08-680-327-3
23	91.5	6.0	1824	4	US-09-228-246-2
24	87	5.7	353	1	US-08-097-831-6
25	86	5.6	432	2	US-08-644-034A-2
26	86	5.6	450	4	US-09-191-879-2
27	84.5	5.5	796	4	US-08-868-699A-2

28	84	5.5	2713	5	PCT-US96-01735-1	Sequence 1, Appl1
29	81.5	5.3	426	2	US-08-644-034A-3	Sequence 3, Appl1
30	78.5	5.1	186	1	US-08-077-256-3	Sequence 3, Appl1
31	78.5	5.1	186	1	US-08-259-672-3	Sequence 3, Appl1
32	78.5	5.1	186	1	US-08-459-351-3	Sequence 3, Appl1
33	78.5	5.1	186	1	US-08-460-353-3	Sequence 3, Appl1
34	78.5	5.1	186	5	PCT-US94-06654-3	Sequence 3, Appl1
35	78.5	5.1	189	3	US-09-209-668-11	Sequence 11, Appl1
36	78.5	5.1	189	3	US-09-356-952-1	Sequence 1, Appl1
37	78	5.1	459	2	US-08-472-172-2	Sequence 2, Appl1
38	76	5.0	1112	2	US-08-714-402-2	Sequence 2, Appl1
39	74.5	4.9	189	2	US-08-429-964-82	Sequence 82, Appl1
40	74	4.8	379	3	US-08-840-204-3	Sequence 3, Appl1
41	74	4.8	390	1	US-08-121-714-6	Sequence 6, Appl1
42	74	4.8	390	1	US-08-477-108A-6	Sequence 6, Appl1
43	74	4.8	390	2	US-08-477-112-6	Sequence 6, Appl1
44	74	4.8	390	5	PCT-US93-08322-6	Sequence 6, Appl1
45	74	4.8	399	1	US-08-565-386-16	Sequence 16, Appl1

ALIGNMENTS

```

RESULT 1
US-09-157-603-1
; Sequence 1, Application US/09157603
; Patent No. 6174694
; GENERAL INFORMATION:
; APPLICANT: Havre, Pamela A.
; APPLICANT: Rice, Michael C.
; APPLICANT: Holloman, William K.
; APPLICANT: Kmiec, Eric B.
; TITLE OF INVENTION: REC2 Kinase
; FILE REFERENCE: 7991-034-999
; CURRENT APPLICATION NUMBER: US/09/157,603
; CURRENT FILING DATE: 1998-09-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-09-157-603-1

Query Match          21.1%; Score 323; DB 4; Length 350;
Best Local Similarity 32.2%; Pred. No. 1.6e-29;
Matches 92; Conservative 44; Mismatches 100; Indels 50; Gaps 7;

QY 27 SPDLSSQKHITTTGSGDINDITLGGICKEVTEIGVGVGKTLGIQLAINVOIPVECGGL 86
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 75 SADFSAPFLSTLISALDELHGVAGCSLTLETGPGGCKTQFCIMMSTLATLPNNMGL 134

QY 87 GGNAYITDEGSPWERYQIAGCIRDTLEHPP--HSHEKSSSVOKOLOPERLP-ADI 142
   | | | | | : : : : : | : : : : : | : : : : : | : : : : : |
Db 135 EGNVYVIDTESAFSARLVEIAE-----SRPRYFTEELLLTSSRVHLYRELTCD 187

QY 143 YYRRCISTYEQIAYIVYMKFLREHKDVRIYIDSTYHFRODFE-----DLATFRVLS 197
   | | | | | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 188 VLQRTISLEETI-----SKGKLVLDVSAYVRKEFDQLOGNKKRKFLTA 236

QY 198 GSILKMKAKTYNLAVALVLLNOYTKF-----TEGSPQTLTL 235
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 237 REASSLKLYALEERSIVILLNQTTLSGALASQADLVSPADLSLSEGTSSSCVIAL 296

QY 236 GDSWSHCNTRLILHNGNERYAHL-DKSPSLPVASAPAVYTGKI 280
   | : | | | | | | : : | : : | : : | : : | : : | : : | : : |
Db 297 GNTWSHVNTRLIQLYDSEKROILAKSPPLAFSTFYVYTIKEEGH 342

RESULT 2
US-09-587-436-1
; Sequence 1, Application US/09587436

```


RESULT 9

US-08-373-134D-2

Sequence 2, Application US/08373134D

Patent No. 5780296

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric

APPLICANT: Holloman, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE

TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/373,134D

CLASSIFICATION: 435

FILING DATE: January 17, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Fiebel, Thomas E.

REGISTRATION NUMBER: 29,258

REFERENCE/DOCKET NUMBER: 7991-007

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 781 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-373-134D-2

Query Match 8.2%; Score 126; DB 1; Length 781;

Best Local Similarity 20.5%; Pred. No. 9.4e-06; Indels 114; Gaps 13;

Matches 62; Conservative 52; Mismatches 74;

Query 28 DELS---OKHI-TTSGDLDNDILGGGICHKEVTEIGVPGVGTQGLQILAIN----- 76

Db 215 DVLISGRORHVFSSGSELDLGGVRSVAVLEIVGESGSGKTQMAIOVCTYAAALGLVP 274

Query 77 -----VOIPEVCG-----GIGGKRAVYTD 94

Db 275 LSQADDDHKGNNTFQSRPFVADPIHASTKDDTLSDILQSYGMEPSIGSHRGK--ACYTT 332

Query 95 TEGS---FMERYVOIAEGCIRDLIEH-FPHSHEKSSVY-----KOLQPER 137

Db 333 SSGERAASHIVRALELASFALINERFDRVYVCDPTQSSQDADGRDALLAKAQQGLRRQ 392

Query 138 FLADIIYFRICSYTEQIAVINMEKFLR-----EHKRVYIIVDSVT 179

Db 393 ALANLHI--AC-----VADVEALEHALKYSPLGLIRRLMSKROSGVSRIGVAVVNDLP 445

Query 180 FHFRRQ-----DFEDLARTVYLSGLSLKMKIA-----KTYNLAVALVLLNOVTT 222

Db 446 ALFQDDQAASDIDSLFORSKMLVEITADALKRISAVQWRGASDGCSSAGRAVVLNHNVS 505

Query 223 KF 224

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

US-09-114-637-2

Sequence 2, Application US/09114637

Patent No. 5945339

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric

APPLICANT: Holloman, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE

TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/114,637

CLASSIFICATION:

FILING DATE:

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER: US/08/373,134

FILING DATE: January 17, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Fiebel, Thomas E.

REGISTRATION NUMBER: 29,258

REFERENCE/DOCKET NUMBER: 7991-007

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 781 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-114-637-2

Query Match 8.2%; Score 126; DB 2; Length 781;

Best Local Similarity 20.5%; Pred. No. 9.4e-06; Indels 114; Gaps 13;

Matches 62; Conservative 52; Mismatches 74;

Query 28 DELS---OKHI-TTSGDLDNDILGGGICHKEVTEIGVPGVGTQGLQILAIN----- 76

Db 215 DVLISGRORHVFSSGSELDLGGVRSVAVLEIVGESGSGKTQMAIOVCTYAAALGLVP 274

Query 77 -----VOIPEVCG-----GIGGKRAVYTD 94

Db 275 LSQADDDHKGNNTFQSRPFVADPIHASTKDDTLSDILQSYGMEPSIGSHRGK--ACYTT 332

Query 95 TEGS---FMERYVOIAEGCIRDLIEH-FPHSHEKSSVY-----KOLQPER 137

Db 333 SSGERAASHIVRALELASFALINERFDRVYVCDPTQSSQDADGRDALLAKAQQGLRRQ 392

Query 138 FLADIIYFRICSYTEQIAVINMEKFLR-----EHKRVYIIVDSVT 179

Db 393 ALANLHI--AC-----VADVEALEHALKYSPLGLIRRLMSKROSGVSRIGVAVVNDLP 445

Query 180 FHFRRQ-----DFEDLARTVYLSGLSLKMKIA-----KTYNLAVALVLLNOVTT 222

Db 446 ALFQDDQAASDIDSLFORSKMLVEITADALKRISAVQWRGASDGCSSAGRAVVLNHNVS 505

Query 223 KF 224

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Query 506 AF 507

Db 506 AF 507

Db 310 QYKLNLIHQRLRKERVDVITIDYIQLINSSVGENRQNEIADI---SRTLRGLASEL 366

QY 204 MKIAKTYNLAVLLNQVYTK 223

Db 367 -----NIPYVCLSLQSLRK 379

RESULT 13
 US-08-468-544-8
 Sequence 8, Application US/08468544
 Patent No. 6248563
 GENERAL INFORMATION:
 APPLICANT: Ratti, Giulio
 APPLICANT: Comanducci, Maurizio
 APPLICANT: Tecce, Mario F.
 APPLICANT: Giuliani, Marzia M.
 TITLE OF INVENTION: pCTD PLASMID ISOLATED FROM CHLAMYDIA
 TITLE OF INVENTION: TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED BY
 TITLE OF INVENTION: THEM; RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID
 TITLE OF INVENTION: GENES IN HETEROLOGOUS SYSTEMS, PREPARATION OF SAID
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
 STREET: 301 N. Washington Street
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22046-0747
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/468,544
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/661,820
 FILING DATE: 28-FEB-1991
 APPLICATION NUMBER: IT MI 91A000314
 FILING DATE: 07-FEB-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Svensson, Leonard R.
 REGISTRATION NUMBER: 30,330
 REFERENCE/DOCKET NUMBER: 1267-202P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-241-1300
 TELEFAX: 703-241-2848
 TELEX: 246345
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 451 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-468-544-8

```

Query Match          7.0%;   Score 107;   DB 4;   Length 451;
Best Local Similarity 25.5%;   Pred. No. 0.00067;
Matches 51;   Conservative 39;   Mismatches 74;   Indels 36;   Gaps 9

QY  37 TGGGDLND---TTGGGHCKEVTEIGGVGVGKCTGTLGTLATVNOVPE--CGGIGGKAV  91
      ||::|||  |||  |  |::|||  |::||:::  ::||  |
Db  203 TGYKIDDDGVLLANG---NFVILAAPRSIGKTLAIDMAINLAVYQRRVGFELS----  254
QY  92 YIDTEGSEFVVERVYQIAECCIRDILFHPHSHKSSSVOKLOPREFLADIVYFRICSYT  151
      ||::|||  |||  |  |::|||  |::||:::  ::||  |
Db  255 -LEMSAGQIVERTIANLTGISGKLOKGLDISKEELFVREAGETVA---ESHFYICSDS  309
QY  152 E-QIAYINNMKFLRHKDKVRIYIID-----SVTFHRRQPEDALRTRVYSGISLKL  203
      ::::|  ||::  |::||  |:::  |  |:::  |  |::||  |

```

```

Db      310  QYKLNLIHQRLRKKEKERVIVIFIDYIQLINSVGENQRQNEIADI---SRTLRLGASSEL 366
Qy      204  MKIAKTYNLAVVLLNQVPTTK 223
           | : : | : | : |
Db      367  -----NIPYVCSLSQLSRK 379

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Query Match Similarity      6.6%; Score 101; DB 1; Length 273;
Best Local Similarity      22.7%; Pred. No. 0.0015;
Matches 47; Conservative   27; Mismatches 67; Indels 66; Gaps 8.

QY      ITTGGSDLNDILG-GGINCKEYTEIGVPGVGKTOGLDIALINVOIPVECGIIGKRAYVI 93
       1-111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      41 ITSGLDLIDLAGIGGVPRGRITIEIYPGESSKTTL-----SLHTIAECOKNGGCAPFI 94
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      DLEGSMVRVYQIAEGCIRLDILEHPPHSHESSVSQKOLQPERFLADLYIF-RICSYTE 152
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      95 DAENAL-----DVYAKRLGVDTTE 113
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      153 QIAY-----INMEKEFLREHKDVRIVIIDSVY----FHFRODFED--LATFTVL 196
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      114 NLLVSPSTGEALELELETITRS-GGIDLVVVDVSVALTPRKAEIDMDQDHVGLGARLM 172
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      197 SGLSLKMKIARTYNLA VVLLNQVTTK 223
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      173 SHALKRITGVLRHKMNTTLTFINOIRKK 199

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 04:00:14 ; Search time 1172.69 Seconds
(without alignments)
11881.662 Million cell updates/sec

Title: US-09-537-654-1

Perfect score: 1474

Sequence: 1 tcgaccacagctgcgcact.....aaaaaaaaaaaaaaaaaaaaa 1474

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_est3.*

4: gb_est4.*

5: gb_est5.*

6: gb_est6.*

7: gb_est7.*

8: gb_est8.*

9: gb_est9.*

10: gb_est10.*

11: gb_est11.*

12: gb_est12.*

13: gb_est13.*

14: gb_est14.*

15: gb_est15.*

16: gb_est16.*

17: gb_est17.*

18: gb_est18.*

19: gb_est19.*

20: gb_est20.*

21: gb_est21.*

22: gb_est22.*

23: gb_est23.*

24: gb_est24.*

25: gb_est25.*

26: gb_est26.*

27: gb_est27.*

28: gb_est28.*

29: gb_est29.*

30: gb_est30.*

31: gb_est31.*

32: gb_est32.*

33: gb_est33.*

34: gb_est34.*

35: gb_est35.*

36: gb_est36.*

37: gb_est37.*

38: gb_est38.*

39: gb_est39.*

40: gb_est40.*

41: gb_est41.*

42: gb_est42.*

43: gb_est43.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_esthum29.*
64: em_esthum30.*
65: em_esthum31.*
66: em_esthum32.*
67: em_esthum33.*
68: em_esthum34.*
69: em_esthum35.*
70: em_esthum36.*
71: em_esthum37.*
72: em_esthum38.*
73: em_esthum39.*
74: em_esthum40.*
75: em_esthum41.*
76: em_esthum42.*
77: em_esthum43.*
78: em_esthum44.*
79: em_esthum45.*
80: em_esthum46.*
81: em_esthum47.*
82: em_esthum48.*
83: em_esthum49.*
84: em_esthum50.*
85: em_esthum51.*
86: em_esthum52.*
87: em_esthum53.*
88: em_esthum54.*
89: em_esthum55.*
90: em_esthum56.*
91: em_esthum57.*
92: em_esthum58.*
93: em_esthum59.*
94: em_esthum60.*
95: em_esthum61.*
96: em_esthum62.*
97: em_esthum63.*
98: em_esthum64.*
99: em_esthum65.*
100: em_esthum66.*
101: em_esthum67.*
102: em_esthum68.*
103: em_esthum69.*
104: em_esthum70.*
105: em_esthum71.*
106: em_esthum72.*
107: em_esthum73.*
108: em_esthum74.*
109: em_esthum75.*
110: em_esthum76.*
111: em_esthum77.*
112: em_esthum78.*
113: em_esthum79.*
114: em_esthum80.*
115: em_esthum81.*
116: em_esthum82.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	% Match	Query Length	DB ID	Description
1	438	29.7	596	173	BG051275
2	382	25.9	543	155	BG605725
3	372.2	25.3	614	136	BG606048
4	367.2	24.9	603	135	BES17872
5	311	21.1	523	251	AZ918279
6	300	20.4	495	251	AZ918278
7	252.8	17.2	481	173	BG050955
8	244.6	16.6	670	114	AN310460
9	177.6	12.0	606	113	AW203630
10	102.4	6.9	506	120	AW84884
11	100.4	6.8	550	150	BG471853
12	99.6	6.8	692	177	AI184177
13	98	6.6	658	167	BE389098
14	98	6.6	683	167	BEA07492
15	98	6.6	778	167	BG177028
16	95.6	6.6	676	154	BG474752
17	95	6.4	669	154	BG506121
18	92.4	6.3	899	168	BF672049
19	91.8	6.2	572	149	BF516288
20	88.8	6.0	634	234	AO848265
21	88.2	6.0	504	114	AM327505
22	85.2	5.8	460	148	BF448483
23	82.6	5.6	527	12	AA861390
24	80.2	5.4	468	31	AV629583
25	78.8	5.3	371	10	AA641847
26	77.4	5.3	545	174	BG313647
27	77.2	5.2	635	153	BG403188
28	77.2	5.2	870	146	BF240482
29	77.2	5.0	787	106	AL533302
30	72.4	4.9	535	154	BG521224
31	70	4.7	606	118	AW604548
32	66.8	4.5	744	165	BE281543
33	66.6	4.5	899	245	AZ544861
34	66.6	4.5	938	248	AZ687391
35	66	4.5	975	206	AL533348
36	65	4.4	900	247	AZ677092
37	65	4.4	925	245	AZ549256
38	64.4	4.4	647	131	BE492333
39	63.4	4.3	649	231	AO641696
40	61.6	4.2	481	101	AA702444
41	60.8	4.1	892	146	BF216102
42	60	4.1	984	219	CNS00554
43	59.8	4.1	480	11	AA779979
44	59	4.0	489	226	AQ254558
45	59	4.0	819	106	AL330472
ALIGNMENTS					
RESULT 1					
BG051275		596 bp	mrna	EST	25-JAN-2001
LOCUS					
DEFINITION	FM1_54.D07.g1.A003 floral-induced Meristem 1 (FM1) Sorghum				
ACCESSION	BG051275				
VERSION	BG051275.1				
KEYWORDS	GI:12504866				
SOURCE	EST.				
ORGANISM	Sorghum proproinquum.				
REFERENCE	Sorghum proproinquum				
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.				
TITLE	1 (bases 1 to 596)				
	Cordeonier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt L.H.				
	AN EST database from Sorghum: floral-induced meristems				

JOURNAL		COMMENT	
Unpublished (2000)		Contact: Cordonnier-Pratt MM	
Department of Botany		The University of Georgia	
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA		Tel: 706 542 1860	
Fax: 706 542 1805		Email: mmpratt@uga.edu	
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.		Seq primer: PolyTmIX	
High quality sequence start: 53		High quality sequence stop: 594	
POLYA=Yes.		Location/Qualifiers	
1. 596		/organism="Sorghum propinquum"	
/db_xref="taxon:132711"		/clone_lib="Floral-Induced Meristem 1 (FMI)"	
/note="Organ: Floral-Induced Meristem; Vector: pBluescript II from Lambda Zap II; Site.1: XhoI; Site.2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."			
BASE COUNT	170 a	132 c	129 g
ORIGIN			165 t
Query Match	29.7%;	Score 438;	DB 173;
Best Local Similarity	92.5%;	Pred. No. 2.8e-105;	Length 596;
Matches 472;	Conservative 0;	Mismatches 35;	Indels 3;
			Gaps 1
QY	910	ttgaagttaatgaagatctgcaagaacatataacttgcagcttgctctgttgaaaccaagtc	969
DB	1	ttgaacttttaattgaagtttcaaaagacatattaaccttgcagcttgctctgttgaaaccaagtc	60
QY	970	actaacaaatttacaagaagtgatcttaacttaattgaccttgcctctaggtgacagctgctgc	1029
DB	61	actactaaattttactgaagggtcattttcaatttaacgtcttgctctaggtgacagctgctgc	120
QY	1030	cactcatgcaagaccggttgattctgcactgtaagtggagaacgaacgacatctt	1089
DB	121	caactctatgcacgaacccgctgattttgtactggaatgggaaacgaacgacgacacactt	180
QY	1090	gattaagtccttctactcttcactgactgactgaacccgttatgacagtgacgaagaagatt	1149
DB	181	gataaactcttcttcttcaacttccagtagcttcacgacctatgacagtgacgaagaagatt	240
QY	1150	agaagatcgttgagctcaaacacacgaacgagccggaagtaacgtagcatcttgggtgtcaa	1209
DB	241	agagatgctgtgagtcacaaacacgaacgagccggaagtaacgtagcatcttgggtgtcaa	300
QY	1210	gcaactgtatgtccactaagctccttcagctcttcttcgcacatgacatctttagtagtag	1269
DB	301	gcactgtgtatctccactaagctccttcagctcttcttcgcacatgacatctttagtagtag	360
QY	1270	agaatgagactgagaatagtagcatcttgggtgttgatcttgcagtgcttggcgttggtcta	1329
DB	361	agatgagactgagaatagtagcatcttgggtgttgatcttggcgttggcgttggtcta	420
QY	1330	ccaaccaaccttaagagagaagtaataacacgaacgaacgagctaatataggtgtttatc	1389
DB	421	caaacaccgacc--atgagagaagtagatgcaacacgaacgagctaatataggtgtttatc	477
QY	1390	tgaacatctggccactcgtatcatcagtaa	1419
DB	478	tgaacatctggccactcgtatcatcagtaa	507

RESULT	2	543 bp	mRNA	EST	17-Apr-2001
LOCUS	BG605725				
DEFINITION	RH122_80_c06_g1_n003 Rhizome2 (RH122)		Sorghum propinquum cDNA, mRNA		
ACCESSION	BG605725				
VERSION	BG605725.1				
KEYWORDS	EST.				
SOURCE	Sorghum propinquum.				
ORGANISM	Sorghum propinquum				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.				
AUTHORS	1 (bases 1 to 543) Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L. H.				
TITLE	An EST database from Sorghum: Sorghum propinquum rhizomes				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmp@atuga.edu Seq primer: PolyTmix High quality sequence start: 39 High quality sequence stop: 542 POLYA-No.				
FEATURES	location/qualifiers				
Source	1..543 /organism="Sorghum propinquum" /db_xref="taxon:132711" /clone_lib="Rhizome2 (RH122)" /note="Organ: Rhizomes; Vector: pBluescript II from lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."				
BASE COUNT	152 a 122 c 119 g 150 t				
ORIGIN					
Query Match	25.9%; Score 382; DB 155; Length 543;				
Best Local Similarity	91.68; Pred. No. 1.9e-90;				
Matches	416; Conservative 0; Mismatches 35; Indels 3; Gaps 1;				
QY	966 agctcctaactaattcaagaaggtcatttcaattgactctgtctagtgtgacactg 1025				
Db	1 AGTACCTACTAAATTTTACTGGAAGGTGCTATTTCANTTAACCTTCTCTAGGTGACACTG 60				
QY	1026 gtccacatcatgcaacgaacgggttgattctgtgacatggaatgggaacgaacatacgcaca 1085				
Db	61 GTCCACACTCAAGCAAGAACCGGCTGATTTTGTACTGGATGGGAACGAACGATCGCACA 120				
QY	1086 tcttgataatgctctcttcaacttccagtagcctgacgacccgfatgcagtgtgaagcaag 1145				
Db	121 CTTGTATAGTCTCTCTTCTTACTTCCAGTAGTGCCTCAGCACCGTATCAGTACAGGCAAAAG 180				
QY	1146 gattagagatgctgtgagctcaaacacacaaacgaacccgagctgaacgtatgacatcttggtg 1205				
Db	181 GATTAGAGATGCTGTGAGTCCAAACCAACGAACGACCGGATACGTACGATTTTGGGGCA 240				
QY	1206 tcaagcaattgtatgctcactaagcctctgacgtcttccttcgcacatgagatctttagact 1265				
Db	241 TCAAGCACTTGTATCTCCACATATGCTCTTGTGACGCTTCTTCACCATGATGATTTGGGGCA 300				
QY	1266 agtagagagacatcggaagaaatagfacatttgttattgttctcagttctgttgccgtg 1325				
Db	301 AGTAGAGGACACTGCGAAGAAATCATACTACTTGTGCTGATTTCTCAATTTGCTTGTGACGCTTA 360				
QY	1326 gctacacacacacattaaagagagaaatataacacacgaacaaagctataatagtgctt 1385				

QY	Db	Accession	Version	Keywords	Source	Organism	Reference Authors	Title	Journal	Comment
QY	1386	tatctgaacatctgcccatacgtaattcagtaa	1419							
Db	418	TATCTGAACATGTGCACATCTTGCATTCAGTAA	451							
RESULT	3									
	BG606048	614 bp	mRNA	EST	17-APR-2001					
	LOCUS	RH1Z2_80.C06.b1_A003	Rhizome2 (RH1Z2)	Sorghum proliquinum cDNA, mRNA						
	DEFINITION	sequence.								
	ACCESSION	BG606048								
	VERSION	BG606048.1	GI:13656019							
	KEYWORDS	EST.								
	SOURCE	Sorghum proliquinum.								
	ORGANISM	Sorghum proliquinum.								
	REFERENCE	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.								
	AUTHORS	Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L. H.								
	TITLE	An EST database from Sorghum: Sorghum proliquinum rhizomes								
	JOURNAL	Unpublished (2000)								
	COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mmpratt@uga.edu Seq primer: JEN REV High quality sequence stop: 605 POLYA-No.								
FEATURES										
Source		location/Qualifiers								
		1..614								
		/organism="Sorghum proliquinum"								
		/db_xref="taxon:132711"								
		/clone_id="Rhizome2 (RH1Z2)"								
		/note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from poly A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."								
BASE COUNT		160 a 146 c 160 g 148 t								
ORIGIN										
Query Match		25.3%; Score 372.2; DB:155; Length 614;								
Best Local Similarity		95.5%; Pred. No. 7.7e-88;								
Matches 383; Conservative		0; Mismatches 18; Indels 0; Gaps 0;								
QY	358	tcgaagagcccaaatgctcgtgatatgtcttcctgatgagctgtcacagaacacatact	417							
Db	214	TCTGGGCCCCGCAANTGCTTGGATATGCTATCTATGACGATACAGAAACACATACT	273							
QY	418	actgtgtctgtgtagcctcaatgacatactgtgtgtggtcggtgattcaccgtcaagaagtact	477							
Db	274	ACTGTGCTGTGTGACCTCAATGACATCTTGTGTGGCGGATTCACGTCAAAAGATTACT	333							
QY	478	gagatcgtgtgcctcccaagggtgtgtgaaactcaactcgtggatcgaactagaatcaat	537							
Db	334	GAGATCGGTGTGTGCCAGGGGTTGTGTAACCTCACTGGGGATTCAACTAGCAATCAAT	393							
QY	538	gtacaaatcccaagtgaatgtgtgtgctgtgtgtggaagcaagttatatagatacagag	597							
Db	394	GTACAAATCCCAAGTGAATATGTGTGGCTTGTGTGGAAAGCAGTTTATTCGATACAGAA	453							
QY	598	ggcagttcatatgtgtgaacgtgtcttaccagaattgtcgaaggtgtattatgaagacatactg	657							
Db	454	GGAAGTTTCATGTTGTAACGTGTCTACCAAGTTTGTCAAGGGTGTATCAGGACATCTAG	513							

QY	658	gagacatttcgcacagccatgagaaatccctctctctcctcccaaacattacagccgtgag	717
DB	514	gagcatttcgcacagccatgagaaatccctctctctcctcccaaacattacagccgtgag	573
QY	718	cgttccctgcgcagatcattcattccgcagatgcagta	758
DB	574	cgttccctgcgcagatcattcattccgcagatgcagta	614
RESULT	4		
BE517872			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			

	Yy	461	actgcaagaagaattctagatcggttgcggtccccaagggttggtcaaaactcaactgggga	520
	Dd	224	ATTGTAAAGAGATTTACTGATAGTAGTGGGCTCCACGAATTGGTAAAATCMACTGGCGCA	283
	Oy	521	ttaacttagaatacattgcfacaatacccagltggaatgtgtgtgacctgtgtlgggaagacg	500
	Dd	284	TTCAGGTAGCAATCAACGTCCAAATTTCCCGTGACTATGTGGCTGTGGCGGCAAAACAA	343
	Oy	581	tttatatatagatcacagggcaggttcacagtgtgtgaacggtgtctaccagattgctggaagyt	640
	Dd	344	TTTTATATTGTGCAGAGGGCACATTTCAATGTTGGAACGGCTATCAATAATGTCTAAAGGT	403
	Oy	641	gtttttggagcatcacacggagggacatttcgccagacacgacgaagaagtcctcttctccaaa	700
	Dd	404	GCATCACTGCATTAATGAGAGTATTTTCCATATGCCACATGACAGAAGTCGATCTGTGCCAAG	463
	Oy	701	aacaatcacacgccttgtagcgttcttcctgtgcygagatatactatacttccggalatgcagtaca	760
	Dd	464	AAATCTGCAGCCCGGAGAGTTCTCTGGCAGGAATCTACTACTTCCGAATATGACGCTACA	523
	Oy	761	cgcgaacaattgcagtcataaactaacatbgaaaaattctctaagagaagataaagtgtgc	820
	Dd	524	CTGAAACGATTTGGGGGATGAATAACTATCTGTGGAGAAATCTCTCGAGAACATTAAGATGGA	583
	Oy	821	gtatagttattatttgtagt 840	
	Dd	584	GAATAGTATTATTATGATAGT 603	
RESULT	5	AZ918279	523 bp DNA	GSS 20-MAR-2001
LOCUS		AZ918279	1006003H03.x2 1006 - Rescuemu Grid G Zea mays genomic, DNA	
DEFINITION			sequence.	
ACCESSION		AZ918279		
VERSION		AZ918279.1	GI:13387563	
KEYWORDS			GSS.	
SOURCE			Zea mays.	
ORGANISM			Zea mays.	
			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
			Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC	
			clade; Panicoideae; Andropogoneae; Zea.	
			1 (bases 1 to 523)	
REFERENCE			Walbot,V.	
AUTHORS			Maze genomic sequences found using engineered Rescuemu transposon	
TITLE			Unpublished (2001)	
JOURNAL			Contact: Walbot V	
COMMENT			Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Possible ligation site of ends cut by 2 different endonucleases. Reverse complemented post-ligation sequence from source sequence. Plate: 1006003 row: H column: 03 Class: transposon-tagged. Location/Qualifiers 1..523 /organism="Zea mays" /cultivar="max background W23/A188/B73" /db_xref="taxon:4577" /clone_lib="I106 - Rescuemu Grid G" /library_type="leaf" /dev_stage="adult" /lab_host="DH10B" /note="Organ: leaf; Vector: Rescuemu (engineered from pBluescript backbone). Site_1: BamHI, Site_2: BglII; Rescuemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescuemu, go to the web	

Db	462	CTACTGCGTCGATTCAGACACTTCAG	487
RESULT	14		
LOCUS	BE407492		
DEFINITION	BE407492	683 bp mRNA	EST
ACCESSION	601300322F1	NIH_MGC_21	Homo sapiens cDNA clone IMAGE:3630461 5', mRNA sequence.
VERSION	BE407492.1	GI:9343942	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 683)		
TITLE	NIH-MGC http://mgc.ncl.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgaps-femail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	cDNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov		
	Plate: LNCM318 row: k column: 06		
	High quality sequence stop.		
FEATURES			
source	Location/Qualifiers		
	1..683		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3630461"		
	/clone_lib="NIH_MGC_21"		
	/russ_type="choriocarcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	195 a 136 c 170 g 182 t		
ORIGIN			
Query Match	6.6%;	Score 98;	DB 167;
Best Local Similarity	60.5%;	Pred. No. 3.7e-15;	Length 683;
Matches 161; Conservative	0;	Mismatches 105;	Indels 0;
		Gaps	0;
QY	392 atgagctgcacagaanaacatcactcactcgtgtcgtgtacccaatgacatcactgtgt	451	
Db	263 AGGAGCATACCCAGGGCTTCATATACCCCTTCGTTCCAGCACTAGATGATTCCTGGGG	322	
QY	452 gcggagatcactgcagaagaagtlactgagatcgtgtggtgtccagggtgtgtaaac	511	
Db	323 GTGGAGTGCCCTTAATGAAACACAGAAATTTGTGTCCACAGGTGTGGAAAAACAC	382	
QY	512 aacttgaggttaactacagaaatcaatgataatcccaatccagttgagttgtgtccctgtgt	571	
Db	383 AATTATGATGCAAGTTGGCAGTAGATGTCAGACATACCAAAATGTTTGGAGGAGTGGCAG	442	
QY	572 ggaagaagcattatataatagatacagaaggtcagttcattgttgaacgtgtcaccagattg	631	
Db	443 GTGCAACAGCTTTTATTATGATACAGAGGGAAGTTTATGCTGATAGAGGTGAGACCTTG	502	
QY	632 ctgaaggtgtatttaggacatctg	657	
Db	503 CTACTGCGTCGATTCAGACACTTCAG	528	

RESULT	15	EST	06-FEB-2001
LOCUS	BG177028	778 bp	mRNA
DEFINITION	602313836F1 NIH_MGC_85	Homo sapiens	CDNA clone IMAGE:4419512 5', mRNA sequence.
ACCESSION	BG177028	GI:12683731	
VERSION	BG177028.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 778)		
TITLE	NIH-MGC http://mhc.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabos-remail.nih.gov Tissue Procurement: Louis Staudt, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10155 row: h column: 09 High quality sequence stop: 710. location/Qualifiers		
FEATURES	1..778		
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4419512" /clone_1ib="NIH_MGC_85" /tissue_type="lymphoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: lymph. Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; cloned unidirectionally; oligo-dT primed. Average insert size 1,867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."		
BASE COUNT	216 a	151 c	200 g 211 t
ORIGIN			
Query Match	6.6%;	Score 98;	DB 174; Length 778;
Best Local Similarity	60.5%;	Pred. No. 3,8e-15;	
Matches 161; Conservative	0;	Mismatches 105;	Indels 0; Gaps 0;
QY	392 atgacgtctcacaagaacacatacctactactggtcttctgtgagcccaatgacatactgtgctg 451		
DB	299 AGGAGCATACCCAGGCGCTTCATATCACTTCCTGTGTTCAGCAGATAGATGATATCTGGGG 358		
QY	452 gcggagatcactgcgaagaagttactgaagatcggctggcgctcccaagggtgtgtaaacctc 511		
DB	359 GGGAGTGCCCTTATATGAACAAACAGAAATTGTGCTGCACACAGCGTGTGAAAAACAC 418		
QY	512 aactggagatcactacagcaatcaatgaacaaatccagcagtggaatgtgtgcttgctggtg 571		
DB	419 AATTATGTATGACAGTTGGCAGTACATGTGTCAGATACCAAGATGTTTGGAGAGAGTGGCAG 478		
QY	572 ggaagaagcttatataagaatacaagaagggcagttcattcattgattgaacgtgtctaccagattg 631		
DB	479 GTGAAGCAGATTTTTATGATACAGAGGGAAGTTTATGTGATGATAGAGTGTGATGACCTTG 538		
QY	632 ctgaagggtgtattataggacatactg 657		
DB	539 CTACTGCTGATTCAGCACCCTTCAg 564		

Search completed: October 2, 2001, 05:13:59

Tue Oct 2 11:16:26 2001

us-09-537-654-1.rst

Page 12

Job time: 4425 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2001, 16:27:27 ; Search time 22.68 Seconds
(without alignments)
1715.065 Million cell updates/sec

Title: US-09-537-654-2
Perfect score: 1530
Sequence: 1 MCDQSGSRNGPQKTYVGAQ.....VTGKGRDAVSSNHRARVT 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mmc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1014	66.3	363	10 Q9FT80	Q9FT80 arabidopsis
2	1010	66.0	363	10 Q9FT75	Q9FT75 arabidopsis
3	988	64.6	332	10 Q22144	Q22144 arabidopsis
4	417	27.3	319	1 Q9YFV1	Q9YFV1 aeropyrum p
5	395	25.8	328	1 Q9YB14	Q9YB14 desulfuroc
6	358	23.4	315	1 Q9UWR5	Q9UWR5 pyrobaculum
7	346	22.6	349	1 Q74036	Q74036 pyrococcus
8	345	22.5	356	1 Q9V233	Q9V233 pyrococcus
9	330.5	21.6	326	5 Q9NCP0	Q9NCP0 giardia lam
10	318	20.8	342	13 Q9DGC3	Q9DGC3 cynops pyr
11	316	20.7	364	5 Q61128	Q61128 leishmania
12	313	20.5	340	4 Q9UH11	Q9UH11 homo sapien
13	309	20.2	357	5 Q44246	Q44246 caenorhabdi
14	303.5	19.8	345	3 Q9P972	Q9P972 coprinus ci
15	302	19.7	341	5 Q01680	Q01680 bombyx mori
16	299.5	19.6	323	1 Q9H768	Q9H768 thermoplas
17	299	19.5	332	10 Q9L110	Q9L110 arabidopsis
18	299	19.5	344	10 P93001	P93001 arabidopsis
19	299	19.5	344	10 Q9M5A2	Q9M5A2 hordeum vul

20	298	19.5	369	1 Q93748	Q93748 cenarchaeum
21	297.5	19.4	322	1 Q73948	Q73948 methanococ
22	296.5	19.4	331	5 Q76341	Q76341 tetrahymena
23	294.5	19.2	331	5 Q9GR43	Q9GR43 hydra atten
24	293.5	19.2	386	1 Q9HMM4	Q9HMM4 halobacteri
25	290	19.0	377	5 Q61127	Q61127 leishmania
26	284.5	18.6	337	13 Q9W628	Q9W628 cynops pyr
27	283.5	18.5	340	10 Q9XED7	Q9XED7 zea mays (m
28	282.5	18.5	353	3 Q74569	Q74569 coprinus ci
29	280.5	18.3	353	3 Q93747	Q93747 halobacteri
30	280	18.3	313	5 Q00847	Q00847 trypanosoma
31	280	18.3	348	5 Q01679	Q01679 bombyx mori
32	280	18.3	341	5 Q77135	Q77135 dirosophila
33	280	18.3	373	5 Q9U6M1	Q9U6M1 trypanosoma
34	276.5	18.1	342	10 P94102	P94102 arabidopsis
35	273.5	17.9	344	3 P78613	P78613 emericella
36	273.5	17.9	348	3 P78579	P78579 emericella
37	271.5	17.7	340	10 Q9XED6	Q9XED6 zea mays (m
38	269.5	17.6	347	3 Q9P956	Q9P956 penicillium
39	268.5	17.5	133	1 Q93762	Q93762 uncultured
40	265.5	17.4	353	3 P87210	P87210 neurospora
41	262.5	17.2	133	1 Q93771	Q93771 uncultured
42	260.5	17.0	354	3 Q9UUL2	Q9UUL2 schizosacch
43	255.5	16.7	529	1 Q58001	Q58001 pyrococcus
44	254	16.6	132	1 Q93758	Q93758 uncultured
45	251.5	16.4	133	1 Q93756	Q93756 sulfolobus

ALIGNMENTS

RESULT 1
ID Q9FT80 PRELIMINARY: PRT: 363 AA.
AC Q9FT80:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE RAD51C PROTEIN.
GN RAD51C.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA;
RA Sland N.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ296174; CAC14294.1; -
SQ SEQUENCE 363 AA; 40214 MW; DC38925F9459F25C CRC64;

Query Match 66.3%; Score 1014; DB 10; Length 363;
Best Local Similarity 64.7%; Pred. No. 2.3e-76;
Matches 187; Conservative 56; Mismatches 40; Indels 6; Gaps 2;
QY 5 SGRNRPQKTYVGAQNMAMDMFSDLSQKHITTTGSGDLDDIIGGHCHEVTEIGVPRV 64
DB 80 NGRS-----LINGARNAMDMHEESILPRITSCSDLDNIIIGGGISCRDVEIGVPR 134
QY 65 GKQLTQIAINVOIVECGGLGKRAVYIDTGESFVEVYQIAEGCIRDLIEHPHSHE 124
DB 135 GKQIQIQLSVNQIPREGGLGKRAIYIDTGESFVEVYQIAEGCIRDLIEHPHSHE 194
QY 125 KSSVQKQLOPERLADIIYFRICSTYEQIAVINMEKFLREHKDRIYIDSVTFHFRQ 184
DB 195 HFGANQVQMKPEIDIEENIFRYFVRCSTYEQIAVINMEKFLREHKDRIYIDSVTFHFRQ 254
QY 185 DFDLALRTIRVLSGLSLKMLAKTAKTNLAVALVNLQYTTFTGSGPQLTALDQSMHSCT 244
DB 255 DYDDLQRTIRVLSGLSLKMLAKTAKTSLAVVLLNQTFTGSGPQLTALDQSMHSCT 314

QY 245 NRLLHNGNERYAHLDKSPSLPVASAPYATGKIRDAVSSNHRARY 293
 DB 315 NRVLTYWNGDERAYIDKSPSLPVASASYTVTSRGLRNS-SSSKRYVM 362

RESULT 2

Q9FT75 PRELIMINARY: PRT: 363 AA.

AC 09FT75; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE RAD5C PROTEIN.

GN RAD5C.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RP [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Sland N.;

RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ299424; CAC14091.1; -

SO SEQUENCE 363 AA; 40213 MW; E21D09BC0A79FA7A CRC64;

Query Match 66.0%; Score 1010; DB 10; Length 363;
 Best Local Similarity 64.4%; Pred. No. 4.9e-76;
 Matches 186; Conservative 57; Mismatches 40; Indels 6; Gaps 2;

QY 5 SGRNGPQKKYVAGNANDMFSDLSQKHTTSGDLNDILGGGHCKEVTEIGVPGV 64
 DB 80 NSGRS-----LINGAKNADMDLHEESLPRTTSCDLDNITGGISCRDVTETIGVPGI 134
 QY 65 GKTLGLOLAIVNIPVECGGLGKAVYIDTEGSEFWEVRYOIAEGCIRDLLEHPPHSE 124
 DB 135 GKQIGIQLSVNVPDREGGGLGKAVYIDTEGSEFWEVRYOIAEGCIRDLLEHPPHSE 194
 QY 125 KSSVQKQLOPEFLADITYFRICSTEDQIVNINMEKLRHKKVRYIIVSYTFHFRQ 184
 DB 195 HFGANQVQKPELIDENIFRYCSTEDQIVNINMEKLRHKKVRYIIVSYTFHFRQ 254
 QY 185 DFEFLALRTFVLSGLKIMKIKATNYLAVLLNQVTFTEGSPQTLALDSDSHSCT 244
 DB 255 DYDDLAGQRTFVLSGLKIMKIKATNYLAVLLNQVTFTEGSPQTLALDSDSHSCT 314
 QY 245 NRLLHNGNERYAHLDKSPSLPVASAPYATGKIRDAVSSNHRARY 293
 DB 315 NRVLTYWNGDERAYIDKSPSLPVASASYTVTSRGLRNS-SSSKRYVM 362

RESULT 3

Q22144 PRELIMINARY: PRT: 332 AA.

AC 022144; 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE PUTATIVE DNA REPAIR PROTEIN (RAD57).
 GN F4L23.21.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RP [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,

RA Venter J.C.;
 RL Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002387; AAB82635.1; -
 DR InterPro: IPR001553; -
 SO SEQUENCE 332 AA; 36864 MW; A7F058BFD54B8A66 CRC64;

Query Match 64.6%; Score 988; DB 10; Length 332;
 Best Local Similarity 66.5%; Pred. No. 2.9e-74;
 Matches 183; Conservative 51; Mismatches 39; Indels 2; Gaps 2;

QY 19 AONAMDFSELQKHTTSGDLNDILGGGHCKEVTEIGVPGVQKQIGIOLAIVNQ 78
 DB 59 AKNAMDMLHEESLPRTTSCDLDNITGGISCRDVTETIGVPGIQTQIGIOLSVNQ 118
 QY 79 IPVCGGLGKAVYIDTEGSEFWEVRYOIAEGCIRDLLEHPPHSESSVQKQLOPER 138
 DB 119 IPREGGGLGKAVYIDTEGSEFWEVRYOIAEGCIRDLLEHPPHSESSVQKQLOPER 178
 QY 139 LADYFTRICSTEDQIVNINMEKLRHKKVRYIIVSYTFHFRQDDELALRTFVLSG 198
 DB 179 LENTFTRVCSYTDQIVNINMEKLRHKKVRYIIVSYTFHFRQDDELALRTFVLSG 237
 QY 199 LSKLTKIAKTYNLAIVLLNQVTFTEGSPQTLALDSDSHSCTNRLTLHNGNERYA 258
 DB 238 MALKFMKIAKTFSLAVLLNQVTFTEGSPQTLALDSDSHSCTNRLTYWNGDERYA 297
 QY 259 HLDKSPSLPVASAPYATGKIRDAVSSNHRARY 293
 DB 298 YIDKSPSLPVASASYTVTSRGLRNS-SSSKRYVM 331

RESULT 4

Q9EYF1 PRELIMINARY: PRT: 319 AA.

AC 09EYF1; 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE 319AA LONG HYPOTHEICAL RADA PROTEIN.

GN APE0119.

OS Aeropyrum pernix.

OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococaceae;

OC Aeropyrum.

OX NCBI_TaxID=56636;

RP [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=K1;

RA MEDLINE-99310339; PubMed-10382966;

RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
 Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankel A., Kosugi H.,
 Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 Yamazaki Y., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1.";
 RT DNA Res. 6:83-101(1999).
 RL EMBL: AP000058; BAA79030.1; -
 DR InterPro: IPR001553; -
 DR InterPro: IPR003593; -
 DR PRINTS: PR00142; RECA.
 DR SMART: SM00382; AAA.1;
 SO SEQUENCE 319 AA; 35320 MW; 714E8859D619E25 CRC64;

Query Match 27.3%; Score 417; DB 1; Length 319;
 Best Local Similarity 36.0%; Pred. No. 7.8e-27;
 Matches 102; Conservative 61; Mismatches 86; Indels 34; Gaps 6;

QY 13 QKVVSGA-----QANAMDFSELQKHTTSGDLNDILGGGHCKEVTEIGVPGV 64
 DB 56 QKIVDAAREALNIDFKVAYDKIESNMKITTGSRNLDLGGIGTETELFGFGS 115

OY	65	GKTOGIGDLAIWQJLPRVCGGLGGAGAVYIDTSGSMVEVRVYOIAGCICRDIIEHPHSHE	124
Db	116	GKTQICHOLSVNVQLPDKDGLLEGAAVVYDTGTFRMERIEDMANGV-----	162
OY	125	KSSSVOKOLOPERFLADIIYYERICSYTQIAVINYMKEFLREHKDVRIIVDSVTFFHQ	184
Db	163	-----GIDPEVKNIYIMIRAINSHHQAIDLVDLFTFMVK-NDNIKLVVDVSYSHFRA	214
OY	185	DF---EDLALPRVLSGLSKLMKTAKEYNLAVALINVOYTTFETGSFQTLALGDS-WIS	240
Db	215	EFPGENIAMROQLLNRLHLQMLMRADIFNAVAVITTNQMARPDEVGYDPQTAVGGHVLG	274
OY	241	HSCNRLLILHMN-GNERVAHLDPSPSLPVASAPYAVTGKIRD	282
Db	275	HAPGRVYLKKSNGKNKRITARVYDAHLEBGETVFATTEMGIRD	317
RESULT	5		
O9Y8JA			
ID	O9Y8JA	PRELIMINARY;	PRT; 328 AA.
OC	O9Y8JA		
DT	01-NOV-1999	(TREMBLrel. 12, Created)	
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)	
DE	01-MAR-2001	(TREMBLrel. 16, last annotation update)	
DE	RECOMBINATION/REPAIR PROTEIN RADA.		
GN	RADA.		
OS	Desulfurococcus amyolyticus.		
OC	Archaea; Crenarchaeota; Desulfurococcales; Desulfurococaceae;		
OC	Desulfurococcus.		
OX	NCHI_TaxID=94694;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Z533;		
RA	kil Y.;		
RL	Submitted (Apr-1999) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; AF145465; AAD33955.1; -		
DR	InterPro; IPR001553; -		
DR	InterPro; IPR003593; -		
DR	PRINTS; PR00142; RECA.		
DR	SMART; SM00382; AAA; 1.		
SQ	SEQUENCE 328 AA; 36434 MM; 7E30A77DE55FC75E CRC64;		
<hr/>			
Query Match 25.8%; Score 395; DB 1: Length 328;			
Best Local Similarity 38.2%; Pred. No. 5.4e-25;			
Matches 97; Conservative 50; Mismatches 81; Indels 26; Gaps			
OY	35	ITTSGSDNLNDLIGGCIHCKEVETIGVGVGVTQGIOLAIVQJLPVCGGLGGAGAVYID	94
Db	93	ITTSKSIDELLGGVEVKRTTEFFEGXGSGTQICHOLSVNVQLTPRKGGINGAAVYID	152
OY	95	TGSGMVERVYQIAEGCTRDILIEHPHSHEKSSSVOKOLOPERFLADIYYERICSYTEOI	154
Db	153	TEGTFMRIRIAMARAL-----GIDPKKVADNTIYMRAYNSDHQI	192
OY	155	AVINYMKEFLREHKDVRIIVDSVTFFHRQD---EDLALPRVLSGLSKLMKTAKEYN	211
Db	193	AIVDELFTFVPEN-DYRLVILDSVSHERRAEIPGREHILAERQCKLNSHLQMLRLAEVN	251
OY	212	LAVVLINOVTTKFTBGSFQTLALGDS-WSHSCTRLLILHMN-GNERVAHLDPSPSLPYA	269
Db	252	VAVVVTNOVMARPDEVGYDPQTAVGCHVLAHNPGRVIDLRKSKGNKRLARVYDAHLEPEG	311
OY	270	SAPYAVTGKGRIDA	283
Db	312	EVVFYTEGIRDS	325
RESULT	6		
O9UMRS			
ID	O9UMRS	PRELIMINARY;	PRT; 315 AA.
OC	O9UMRS;		

Query	Subject	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
QY	21 NAMDSEDELSQKHITTSQSGDLNDLIGGICHCETETGCGYCGKQQLGLOLAIVNQIP	23.4%	358	94	53	93	28	6
Db	70 SALEVYERRKRIIRISTGVRALDELGGIETRAVTEVAGGEGSKQLCHOLAVMYOLA	Best Local Similarity 35.1%; Pred. NO. 6.1e-22;						
QY	81 VECGLGKGRKVVYIDTSGFMVERVQIAGCIRDLIEHPPSHKSSSVOKOLPERFLA							
Db	130 EERGGGLGKAKIYIDTENTPRPERIQIACA							
QY	141 DIIYFRICSTQEOIAVINMKEFLREHKDVRILVIIDSYTFHRRDF--EDLALRTFVLS							
Db	170 NIFYARAYSSDHQMLIDQAKSTIRQH-VNALLIADVSIAHFSEFGRENLAEROOKIN							
QY	198 GLSLKLMKIAKTYNLAVLLINOVYTKFTBGSF-QLTALDSDMSHSCTNRLILHMN-GN							
Db	229 KHAVADLRLDADVAIVVITNQVMAQ-PDVEFGNPLRPAGGNILAHATYRMLRKSKE							
QY	255 ERYAHLDRKSPSLPVASAPYAVTGKGIARD							
Db	288 IRIAKITFDSPYHPEGVSEFRITEBGLVD							
RESULT	7							
ID	074036	PRELIMINARY:	PRT:	349	AA.			
NC	074036							
DT	01-NOV-1998 (TReMBLrel. 08, Created)							
DT	01-NOV-1998 (TReMBLrel. 08, Last sequence update)							
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)							
DE	RECOMBINASE.							
GN	RADA.							
OS	Pyrococcus furiosus.							
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.							
OX	NCBI_TaxID=2261;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN-DSM 3638;							
RA	Diruggiero J., Brown J.R., Bogert A.P., Robb F.T.;							
RT	"DNA repair systems in Archaea: Mementos from the last universal							
RT	common ancestor.";							
RL	J. Mol. Evol. 0:0-0(1998).							
DR	EMBL: AF052597; AAC3498.1; -.							
DR	HSSP: P03017; ZREB.							
DR	InterPro: IPR000445; -.							
DR	SMART: SM00382; AAA; 1.							
SO	SEQUENCE 315 AA; 3518 MW; 3EPD424DBCA5F8 CRC64;							

OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae; Cynops.
 OX NCBI_TaxID=8330;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=TESTIS;
 RA Yezawa T., Yamamoto T., Nakayama Y., Hamada S., Abe S.;
 "Conversion from mitosis to meiosis: morphology and expression of PCNA
 and DMC1 during newt spermatogenesis."
 RT Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AB041944; BAB16892.1;
 DR SEQUENCE 342 AA; 37981 MW; 868E002AEAF0BEEB CRC64;

Query Match 20.8%; Score 318; DB 13; Length 342;
 Best Local Similarity 33.5%; Pred. No. 1,4e-18;
 Matches 90; Conservative 44; Mismatches 101; Indels 34; Gaps 8;

OY 29 ELGSK-----HTTGGDNDLILGGHCKEVTETIGVPGVGTOLGIDAINVQIPVEG 83
 DB 92 EYSEKRMVPHITGSGDFKLGGLGIESMAITTFEFPRTGKTQLSHLTCTVATQLEPTD 151
 OY 84 GGLGKAVYIDTEGSMFVERVYQIAEGCIRDILEHFPHSHEKSSSVOKOLOPERFLADY 143
 DB 152 GYTGKVIPTENTFRDR-----LRDIADRFSDH-----DAVIDNL 191
 OY 144 YFRICSTEQIAVINM-EKFLREKDVRIYIDSVFHRQPE---DLALTRVLSGL 199
 DB 192 YARAVTSEHOMELDYAAKHEBAGIFKLIVDSIMALFRVDSGGLAEARQKLAQM 251
 OY 200 SLKLMKIAKTYNLAIVLLNOVTT-KTEGSPQLTLAL---GDSWSHSCNRLILH-WNGN 254
 DB 222 LARLOKISEEVNAVFTYNTAPDPAAGAMSQADPKPKIGIHILAHSTRISLRKRG 311
 OY 225 ERYAHLDSKPSLPVAPAVATGKIRDA 283
 DB 312 LRIAKIVDSPMPENEAFTATAGISDA 340

RESULT 11

O61128 PRELIMINARY; PRT; 364 AA.
 ID O61128
 AC O61128
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE DMC1 HOMOLOG.
 GN DMC1.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA McKeon P.G., Benson F.E.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF062380; AAC16335.1;
 DR HSSP: P03017; 2REB.
 DR InterPro: IPR001553;
 DR InterPro: IPR003593;
 DR SMART: SM00382; AAA; 1.
 SQ SEQUENCE 364 AA; 39527 MW; A5475644CE8AFBE5 CRC64;

Query Match 20.7%; Score 316; DB 5; Length 364;
 Best Local Similarity 34.5%; Pred. No. 2.3e-18;
 Matches 90; Conservative 49; Mismatches 88; Indels 34; Gaps 8;

OY 35 ITTGGDNDLIL-GGHIHCKEVTETIGVPGVGTOLGIDAINVQIPVEGGLGKAVYI 93
 DB 126 ISTGSAIDQLGGGIERSTTEAFGERPTKTOIGHTLCVTCOLPLEMGKNAKAVY 185
 OY 94 DTGGSFVERVYQIAEGCIRDILEHFPHSHEKSSSVOKOLOPERFLADYIFRICSYEQ 153

DB 186 DTGTFPRPRIRPIAREFGMD-----SNSV-----LDNIIYARATYTHEQ 225
 OY 154 IAVINMEFLAEKHDVRIYIDSVFHRQPE---DLALTRVLSGLSKIMKIATY 210
 DB 226 AHLNVAKMAEDQ-FSLVVDSTALFRVDSGGLAEARQKLAQLSOLITIAEEF 284
 OY 211 NLAIVLLNOVTKRTFGSQTFLA-----LGDMSHSCNRLILH-WNGENRYAHLDS 263
 DB 285 NIAVITTYNOVVS--DEGASMEVADPKPKVGGHILAHSTTSLRKRGDQRCVKIDS 342
 OY 264 PSLPVASAPYATGKIRDAV 284
 DB 343 PSLPELECYTISEGQIIDAV 363

RESULT 12

O90H11 PRELIMINARY; PRT; 340 AA.
 ID O90H11
 AC O90H11
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE DJ199H16.1 (DMC1 (DOSAGE SUPPRESSOR OF MCK1, YEAST HOMOLOG) MEIOSIS-
 DE SPECIFIC HOMOLOGOUS RECOMBINATION PROTEIN (LIM15)))
 GN DMC1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Blakey S.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF022320; CAB45656.1;
 DR HSSP: P03017; 2REB.
 DR InterPro: IPR001553;
 DR InterPro: IPR003593;
 DR SMART: SM00382; AAA; 1.
 SQ SEQUENCE 340 AA; 37681 MW; 040A6E4CF1EBFA2 CRC64;

Query Match 20.5%; Score 313; DB 4; Length 340;
 Best Local Similarity 32.7%; Pred. No. 3.7e-18;
 Matches 89; Conservative 45; Mismatches 98; Indels 40; Gaps 8;

OY 29 ELGSK-----HTTGGDNDLILGGHCKEVTETIGVPGVGTOLGIDAINVQIPVEG 83
 DB 90 EYSEKRMVPHITGSGDFKLGGLGIESMAITTEAFGERPTGKTQLSHLTCTVATQLEPTD 149
 OY 84 GGLGKAVYIDTEGSMFVERVYQIAEGCIRDILEHFPHSHEKSSSVOKOLOPERFLADY 143
 DB 150 GYTGKVIPTENTFRDR-----LRDIADRFSDH-----DAVIDNL 189
 OY 144 YFRICSTEQIAVINM-EKFLREKDVRIYIDSVFHRQPE---DLALTRVLSGL 199
 DB 190 YARAVTSEHOMELDYAAKHEBAGIFKLIVDSIMALFRVDSGGLAEARQKLAQM 249
 OY 200 SLKLMKIAKTYNLAIVLLNOVTKRTFGSQTFLA-----GDSWSHSCNRLILH-W 251
 DB 250 LSRLOKISEEVNAVFTYNTAPDPAAGAMSQADPKPKIGIHILAHSTRISLRKRG 306
 OY 252 NGNERVAHLDSKPSLPVAPAVATGKIRDA 283
 DB 307 RGLRIAKIVDSPMPENEAFTATAGISDA 338

RESULT 13

O44246 PRELIMINARY; PRT; 357 AA.
 ID O44246
 AC O44246
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE RECA/RAD51/DMC1-LIKE PROTEIN.
 GN RAD51 OR RDH-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Higashitani A., Sato S.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CB1392;
 RX MEDLINE=99077299; PubMed=9862483;
 RA Rinaldo C., Ederle S., Rocco V., La Volpe A.;
 RT "The Caenorhabditis elegans Rad51 homolog is transcribed into two
 RT alternative mRNAs potentially encoding proteins of different sizes.";
 RL Mol. Gen. Genet. 260:289-294(1998).
 RN (3)
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN (4)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; AB011382; BAA24982.1; -;
 DR EMBL; AF061201; AAD10194.1; -;
 DR EMBL; AL023838; CAB61038.1; -;
 DR InterPro; IPR000445; -;
 DR InterPro; IPR001553; -;
 DR InterPro; IPR003583; -;
 DR Pfam; PF00633; HHH; 1.
 DR SMART; SM00278; Hhhl; 1.
 SQ SEQUENCE 357 AA; 39193 MW; ECEB8C7021204BED CRC64;

Query Match 20.2%; Score 309; DB 5; Length 357;
 Best Local Similarity 33.7%; Pred. No. 8,5e-18;
 Matches 87; Conservative 36; Mismatches 105; Indels 30; Gaps 6;

QY 35 ITTGSGDLNLLGGCHCKEYETIGVPGVGTQGLGQLAINVQIPVECGGLGKAVYID 94
 DB 118 IRTGASLDRLLGGIETGITEYGEYRTGTQLCHSLAVLCQLPDMGGGCKMYID 177
 QY 95 TEGSFNVERVYQIAEGCIKRDILIEHPHSHSEKSSVQKQLQPEF-----LADYIFRI 154
 DB 178 TNAIFRPERIADQ-----RYMDSAHVLENTAVARAYNSEHLM 217
 QY 155 AVINMEKFLREHKDVRIVIDSYTFHRODFE--DLALRTVLSGLSLMKIAKTYN 211
 DB 218 ALIIRAGAMMSER-YAVVIVDCATFAHFRNYTGKGLDAEROMLSAFLKCLATLADEYS 276
 QY 212 LAVVLLNVOYTKFTTGS--FQITIAL---GDSMSCHCNRLILH-WGNENYRAHLKSPS 265
 DB 277 VAVIITNOVAVQVGDGSMFQADAKKPIGHIILAHMSTRLYLKKGGENVAKVAVOSPN 336
 QY 266 LPVASAPYAVTGKIRDA 283
 DB 337 LPEAATATYSITNHGIEDA 354

RESULT 14
 Q9P972 PRELIMINARY; PRT; 345 AA.
 AC Q9P972;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE LIM15/DMC1 HOMOLOG PROTEIN.

OS Coprinus cinereus (Inky cap fungus).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Agaricales;
 OC Coprinaceae; Coprinus.
 OX NCBI_Taxid=5346;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Nara T., Sakaguchi K.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2092483; PubMed=10628861;
 RA Nara T., Saka T., Sawado T., Takase H., Ito Y., Hotta Y.,
 RA Sakaguchi K.;
 RT "Isolation of a LIM15/DMC1 homolog from the basidiomycete Coprinus
 RT cinereus and its expression in relation to meiotic chromosome
 RT pairing";
 RL Mol. Gen. Genet. 262:781-789(1999).
 DR EMBL; AB036801; BAA89533.1; -;
 DR InterPro; IPR001553; -;
 DR InterPro; IPR003593; -;
 DR PRINTS; PR00142; RECA.
 DR SMART; SM00382; AAA; 1.
 SQ SEQUENCE 345 AA; 37606 MW; AB1C15B796F6F0EB CRC64;

Query Match 19.8%; Score 303.5; DB 3; Length 345;
 Best Local Similarity 32.1%; Pred. No. 2,3e-17;
 Matches 85; Conservative 40; Mismatches 95; Indels 45; Gaps 6;

QY 35 ITTGSGDLNLLGGCHCKEYETIGVPGVGTQGLGQLAINVQIPVECGGLGKAVYID 94
 DB 108 ISTGSKLVDDILGGVMSQSITVEYGEYRTGTQLAHMYSVAQLPPEYGAAGKAVYID 167
 QY 95 TEGSFNVERVYQIAEGCIKRDILIEHPHSHSEKSSVQKQLQPEF-----LADYIFRI 147
 DB 168 TEGFRPRDRIRATA-----DRFGVDGTMALFNILYARA 200
 QY 148 CSTYEOIAVINMEKFLREHKDVRIVIDSYTFHRODFE--DLALRTVLSGLSLMK 204
 DB 201 FNSHQWELINECSARFAEDKDERLLIDISIMALFRVYSGRGEISEPOKLAQMLSKLT 260
 QY 205 KIKKTYLAVVLLNOV-----TTKTEGFSQTLALGDSMSCHCNRLILH-WGNENY 257
 DB 261 KLSSEFNIAVLMTNVOVSDGATTFVAGALKPIG-GHILSHASATRIFLKGRAERY 319
 QY 258 AHDKSPSLPVASAPYAVTGKIRDA 282
 DB 320 AKLVDSPPDRPESASRYKLDECGMAD 344

RESULT 15
 O01680 PRELIMINARY; PRT; 341 AA.
 AC O01680;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE DMC1 HOMOLOG.
 GN DMC1.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Bombycoidea; Bombycidae; Bombyx.
 OX NCBI_Taxid=7091;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Larsen J.J.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U94994; AAB53331.1; -;
 DR HSSP; P03017; 2REB.
 DR InterPro; IPR001553; -;
 DR InterPro; IPR003593; -;

Query Match 19.7%; Score 302; DB 5; Length 341;
Best Local Similarity 32.3%; Pred. No. 3.1e-17;
Matches 92; Conservative 49; Mismatches 106; Indels 38; Gaps 10;

[illegible]

Search completed: October 1, 2001, 16:29:07
Job time: 100 sec

